HOGgles: Visualising Object Detection Features

Summary written by Sanika Phatak
February, 09, 2020

Summary: The paper introduces a tool to visualize the feature space as seen by object detectors to analyse the false alarms of detectors better (this is also called feature inversion). The paper suggests that feature space used is heavily responsible for the false positives given by the object detector and hence bigger datasets and better learning algorithms might not be a solution to them.

Related work: The initial work in this domain was done in the paper [4] which first proposed image reconstruction given the SIFT descriptors. All the existing methods have feature specific solutions for visualization the method proposed in this paper is feature independent. The paper [2] inspired benchmarking of the algorithm using humans debugging of object detectors.

Approach: The feature visualization problem is treated as a feature inversion problem for which the paper states three baselines: (i) Exemplar LDA (ELDA) - ELDA detector is trained for $w = \Sigma^{-}1(y - \mu)$ where y is a HOG feature. The inverse is the average of top K detections. (ii) Ridge Regression - Random variables for a grey-scale image and its HOG detector are defined to be normally distributed on a Gaussian and to find its inverse the most likely image is found from the conditional Gaussian distribution. Assuming a closed form Gaussian, the inversion is done by a single matrix multiplication which makes the process fast.(iii) Direct Optimisation - Only the images spanning the natural basis are considered and their first K eigen vectors are found and any image x can be encoded coefficients ρ from set of K by $x = U\rho$ where U represents the set of natural basis. The Paired Dictionary Learning algorithm is introduced as the main feature inversion algorithm. Here, x is the image and y is its HOG descriptor which are written in terms of bases as - $U \epsilon \mathbb{R}^{DxK}$ and $V \epsilon \mathbb{R}^{dxK}$ respectively with shared coefficients $\alpha \epsilon \mathbb{R}^{K}$. HOG features y are first projected on HOG basis V and then the α is projected on the natural basis U.

$$x = U\alpha \; and \; y = V\alpha$$

. The Paired Dictionary Learning problem solution used is inspired from the [3] which contributes to super resolution sparse coding after which the concatenated dictionaries are optimised using SPAMS [1].

Experiments and results: ELDA and Ridge Regression produce blurred images and Direct optimization produces extra noise. Paired dictionary learning tends to produce best visualizations for HOG detectors and is able to recover partially correct colours from HOG descriptors. To benchmark each inversion, it is compared against the original image with normalised correlation where Paired dictionary gives best results with a mean of 0.637. For visualisation benchmark, 2000 windows were sampled from PASCAL VOC 2011 objects and human participants were shown inversions and asked to classify it in 20 categories. Paired dictionary and direct optimisation gave the best visualisations from this experiment. The experiment also suggests that human accuracy on inversions and object detector algorithms are correlated with Spearman's coefficient of 0.77.

Strengths: The paper provides a new insight on tackling the false alarms issue in object detectors. The concept of feature space visualisation and inversion makes it easy for humans to analyse the feature space to come up with better solutions.The existing methods are restricted to specific features but this algorithm is feature independent. The paired dictionary algorithm used for inversion is fast, recovers high frequencies without noise and is partially able to recover colour from the HOG descriptors. The human as well as automatic bench-marking used to evaluate the visualisations covers a wholesome approach.

Weaknesses: The paper provides good visualisations but there is a large scope for improvement. When compared against a greedy reconstruction algorithm which is executed for a long time paired dictionary performs poorly. Also, while inversions there is a loss of information specially of higher frequency. The inversions are highly sensitive to the dimentionality of the data and the performance degrades with decrease in the size of the template.

Reflections: HOGgles introduces a very unique way of looking at false alarms issue in object detection and is a segway for a new research domain for improving feature space. The paper highlights the limitations of the HOG detector and gives a direction to improve its performance. The benchmarking using automatic technique as well as human debugging is a novel idea which should be considered while evaluating visualisations in object detectors. The paper also introduces the concept of object detectors visualising the feature space like humans do and gives a different viewpoint to improve performance apart from increase in training dataset or developing better training algorithms.

References

[1] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. 2009.

[2] D. Parikh and C. Zitnick. A combined corner and edge detectorhuman debugging of machines. 2011.

[3] S. Wang, L. Zhang, Y. Liang, and Q. Pan. Semicoupled dictionary learning with applications to image super-resolution and photosketch synthesis.

[4] H. Weinzaepfel, H. Jegou, and P. Perez. Reconstructing an image from its local descriptors. 2011.