HOGgles: Visualizing Object Detection Features

Summary written by Sahana Srihari
February 10, 2020

Summary: This paper deals with exploring algorithms utilized to visualize feature spaces used for object detection tasks.The paper stresses on the construction of a better defined feature space in contrast to improving learning algorithms or increasing the dataset as they have identified that the high number of false alarms in object detection paradigms is engendered due to poorly defined feature spaces. Main contribution - algorithms to visualize feature spaces for object detection in terms of 4 algorithms to invert object detection features to natural images which is validated via auto benchmark and human studies.

Related work: The author elucidates the extension of work from several sources in regards to how previously image feature inversion and reconstruction took place.The work of [5] and [8] is highlighted for image inversion and reconstruction for SIFT descriptors. [5] [1][8] is mentioned for reconstructing images from SIFT, LBP, and gist features but these features are tied to a specific use case unlike the author's approach to develop independent features. The main contributions seen in this paper is from [6] where they have adopted their method for human debugging of object detectors.[3] for error detection, [2] for impacting component analysis, [7] for image identification closely resembling HOG descriptors,[4] and [9] for picking image regions contributing towards classifier confidence and the associated risk respectfully.

Approach: The methods can be divided into 4 categories for feature visualization:

(i)Baseline A: Exemplar LDA - This is based on the fact that we can extract observations from false positives. The feature y being inverted is used to train an Exemplar LDA and a score w is obtained which is computed across every sliding window. The HOG inverse is then obtained by averaging the top 'k' detections of the image. But this approach is viewed as computationally expensive due to the fact that the sliding window must parse through entire large data sets.

(ii)Baseline B: Ridge Regression - This method follows ridge regression of Gray scale images, where the image X and its corresponding HOG point is normally distributed over a Gaussian. Inversion of this feature y is obtained by deriving the conditional probability of the most likely image occurrence. This is a very fast parametric inversion but is susceptible to picking up only low frequencies.

(iii)Baseline C: Direct Optimization - This method deals with images that only span the natural image basis in order to match the original descriptor. The significant "K" Eigen vectors form the space for the coefficients which are used to represent an image x. Coordinate descent is used to optimise the coefficients.

(iv)Algorithm D: This is the main algorithm the paper highlights. We take an image x and its descriptor $\mathbf{y}$ such that it is represented in terms of a bases U and V with shared α coefficients respectively. y is projected onto V and α is project to U this provides a paired dictionary based on the commonality of α. This problem is optimised using SPAMS.

Strengths: The author proposed a very important solution for the field of computer science by giving us a tool to gain insight onto object detectors fail cases. This was a bleeding edge solution that nobody else had done before. Practically, the author provides 4 efficient solution to solve the underlying problem out of which Paired Dictionary Learning is the most efficient and fastest. Several experiments were executed with these algorithms and backed with experimental proof. The introduction of this concept of visualizing object detection features allowed engineers and scientists to look at an image as how a system would and find the root cause as to why the system is spitting out false negatives. The paper is well structured, easily understandable and sheds light on an important finding that beyond a point of good learning algorithms and datasets there is a strong need of efficient feature picking.

Weaknesses: This algorithm provides 4 different solutions to the problem and then goes on to conclude that one of their algorithms is far better than the others but this algorithm does not compare to the speed of a greedy approach. Although the greedy approach would be computationally expensive, provided the advance in technology, more research should be done in analysing better and faster algorithms on better hardware. This is a minor problem in computer vision and the author provides solutions to it albeit provides pathways for further research.

Reflections: This paper provides multiple algorithms to visualize object detection features. The author admits that his approach is not the most efficient method and more research could be done to find that perfect algorithm. Although this was a good start to answer some of the outlying problems in the computer vision community, this project can be branched out to answer similar questions in the other sub-branches of computer vision. This works very well as this method is feature invariant i.e., it works with any detector which uses any feature.

References

[1] E. d'Angelo, A. Alahi, and P. Vandergheynst. Beyond bits: Reconstructing images from local binary descriptors. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 935–938. IEEE, 2012.

[2] S. K. Divvala, A. A. Efros, and M. Hebert. How important are "deformable parts" in the deformable parts model? In *European Conference on Computer Vision*, pages 31–40. Springer, 2012.

[3] D. Hoiem, Y. Chodpathumwan, and Q. Dai. Diagnosing error in object detectors. In *European conference on computer vision*, pages 340–353. Springer, 2012.

[4] L. Liu and L. Wang. What has my classifier learned? visualizing the classification rules of bag-of-feature model by support region detection. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593. IEEE, 2012.

[5] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

[6] D. Parikh and C. L. Zitnick. The role of features, algorithms and data in visual recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2328–2335. IEEE, 2010.

[7] A. Tatu, F. Lauze, M. Nielsen, and B. Kimia. Exploring the representation capabilities of the hog descriptor. In *2011 IEEE international conference on computer vision workshops (ICCV Workshops)*, pages 1410–1417. IEEE, 2011.

[8] P. Weinzaepfel, H. Jégou, and P. Pérez. Reconstructing an image from its local descriptors. In *CVPR 2011*, pages 337–344. IEEE, 2011.

[9] X. Zhu, C. Vondrick, D. Ramanan, and C. C. Fowlkes. Do we need more training data or better models for object detection?. In *BMVC*, volume 3. Citeseer, 2012.