NeurAda: Combining artificial neural network and Adaboost for accurate object detection

SABER SHAKERI
Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
and
MEHRAN ASHOURAEI
Young Researchers and Elite Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Object detection is a very important technique in computer vision which is mainly used in many applications. Many papers have addressed this problem and proposed different methods to improve the accuracy of detectors. The main disadvantages of common methods in object detection are high time complexity, wrong object detection, not detecting objects. Extracting features and classification are two step of detecting objects. In this paper, a new method is presented to improve some of the disadvantages using Histograms of Oriented Gradient (HOG) as feature extractor and artificial neural network combined with Adaboost (NeurAda) as a classifier to cover weak points of previous works. To evaluate the proposed method, NeurAda was compared to the three top obtained results of Pascal VOC 2011 methods in three categories. NeurAda improved car detection by 8.6%, bicycle detection by 0.8% and pedestrian detection by 5.2% in comparison to best results of Pascal VOC 2011.

Keywords: Artificial Neural Network, Adaboost, Histograms of Oriented Gradient, NeurAda, Classification.

1. INTRODUCTION

Object detection is a very important technique in computer vision which is mainly used in many applications like military, satellite image mining, medical and etc. [Santhi and Bhaskaran 2014]. A common paradigm to address this is to train object detectors which operate on a sub-image and apply these detectors in an exhaustive manner across all locations and scales [Erhan et al. 2014]. Extraction of features is the first step of object detection. The feature extraction step is crucial since the features are going to represent images and, therefore, will be used as input for machine learning or content-based retrieval methods [Ponti et al. 2015]. While extracting, one needs to choose from a wide variety of methods that will work well in some situations, while fail in others [Ponti et al. 2015; Cortes and V. 1995]. Second step is classification. Classification is the arrangement of objects according to their observed similarities. The classifications used in this paper are neural network and Adaboost. An artificial neural network (ANN) is an information processing system which has certain performance characteristics similar to biological neural networks [Hussain et al. 2015]. Adaboost, one of the first practical boosting methods, was introduced by Freund and Schapire in which weak learners are combined and boosted to improve ensemble accuracy and produce a strong classifier where the training data is iteratively sampled, with replacement, to train the weak learner [Kummer and Najjaran 2014]. The main motivation in this study is to develop and analyze the combination of artificial neural network and Adaboost (NeurAda) to obtain a object detector with high accuracy as both ANN and Adaboost already exist but they were not combined together to improve object detection before. The structure of this paper is organized as follows: In section 2, some background and related works are reviewed. Section 3 contains of the basic idea of the proposed method for object detection. In section 4, implementation detail and some experimental results are given. Finally, section 5 includes the conclusion of this study and suggests some future works.

2. BACKGROUND AND RELATED WORK

[Santhi and Bhaskaran 2014] proposed an object detector based on data clustering methods in region based segmentation and shape feature. Most of previous researches have used the k-means and fuzzy k-means for clustering and this paper used the SVM and Adaboost classifiers for object classification. In the proposed method, each cluster needs its own centric and distance calculation for clustering. The main disadvantage of this technique is distance calculation between the pixels which does not produce efficient result in clustering. In classification, SVM classifier needs more parameters for increasing the efficiency and Adaboost is more noise sensitive. To avoid these drawbacks, the authors used region based segmentation with non-Euclidean distance measure for clustering and combined Fisher SVM with modified Adaboost algorithm for object classification. The results showed the region based segmentation and object classification of an image with an increased efficiency in performance analysis graph.

[Cheng et al. 2015] proposed a method of training mixture of weighted SVM models using EM algorithm by the idea of divide-and-conquer approach and discriminatively trained SVM model for object detection. In the paper, they introduced a new part weighted SVM with logistic function to convert its prediction score into pseudo-probability. The part weight is computed by an energy estimation method to reflect the discriminative power of different object parts, and the conversion of prediction score to probability enables the input to be assigned to a proper SVM based on unbiased prediction scores among multiple SVM models. The two modifications fit the joint training process of multiple SVMs into the EM framework, where they iteratively reassigned the object examples into different sub-regions of the entire input space, and then retrained the SVM models corresponding to that sub-region. In that way, the mixture of SVM models became a set of experts to form the mixture of DPMs. Experimental results showed that the proposed method made noticeable improvements over the baseline method, which demonstrated the advantage of the proposed method for training MDPM based models for object detection.

[Kong and Hong 2015] proposed coupled strong classifiers (CSCs) which consist of multiple strong classifiers connected in parallel to achieve further improvement in the Adaboost framework as a strong classifier consists of weak classifiers connected sequentially and usually the detection performance of the strong classifier can be improved increasing the number of used weak classifiers. Complementarity between the classifiers is considered for reducing intra- and inter-classifier correlations of exponential loss of weak classifiers in the training phase, and dynamic programming is used during the testing phase to compute efficiently the final object score for the coupled classifiers. In addition to CSC concept, they also proposed using Aggregated Channel Comparison Features (ACCFs) that take the difference of feature values of Aggregated Channel Features (ACFs), enabling significant performance improvement. To show the effectiveness of their CSC concept, they applied their algorithm to pedestrian detection. Experiments were conducted using four well-known benchmark datasets based on ACFs, ACCFs, and Locally Decorrelated Channel Features (LDCFs). The experimental results showed that their CSCs gave better performance than the conventional single strong classifier for all cases of ACFs, ACCFs, and LDCFs.

[Rios-Cabrera and Tuytelaars 2014] presented a novel template-based approach for fast object detection and investigated the use of Dominant Orientation Templates (DOT) which is a binary template representation as a means for fast detection of objects. They learnt a binary mask for each template that allowed to remove background clutter including relevant context information at the same time. The mask templates then were served as weak classifiers in an Adaboost framework. They demonstrated the proposed method on detection of shape-oriented object classes as well as multi-view vehicle detection and obtained a fast yet highly accurate method for category level detection. They showed how to efficiently transfer meta-data using the top most similar activated templates. Finally, they proposed an optimization scheme for detection of specific objects using the proposed masks trained by the SVM which resulted in an increment of up to 17% in performance of the DOT method without sacrificing testing speed and ability to run the training on real time.

[Tana et al. 2014] proposed an approach to select compact and effective features of a holistic filter and several part filters in a relatively high-dimensional feature space by learning a sparse deformable part model using L1-norm latent SVM. A stochastic truncated sub-gradient descent method was presented to solve the L1-norm latent SVM problem and convergence of the algorithm was proved. Extensive experiments were conducted on the INRIA and PASCAL VOC 2007 datasets. A highly compact feature in the method could reach the state-of-the-art performance. The feature dimensionality was reduced to 12% of the original one in the INRIA dataset and less than 30% in most categories of PASCAL VOC 2007 dataset. Compared with the features used in L2-norm latent SVM, the average precisions (AP) had almost no drop using the reduced feature. With the proposed method, the speed of the detection score computation is faster than that of the L2-norm latent SVM method by 3 times.

[Chitrakar and Huang 2014] proposed a half-partition strategy of selecting and retaining non-support vectors of classification named as Candidate Support Vectors (CSV). The authors also designed an algorithm named Candidate Support Vector based Incremental SVM (CSV-ISVM) that implemented the proposed strategy and materialized the whole process of incremental SVM classification. This work also suggested modifications to the previously proposed concentric-ring method and reserved set strategy. Performance of the proposed method was evaluated with experiments and also by comparing it with other ISVM techniques. Experimental results and performance analyses showed that the proposed algorithm CSV-ISVM was better than general ISVM classifications for real-time network intrusion detection.

[Abdiansah and Wardoyo 2015] proposed results of their work related to complexity analysis of Support Vector Machines (SVM). Their work has focused on SVM algorithm and its implementation in LibSVM. Support Vector Machines is one of machine learning methods that can be used to perform classification task. Many researchers using SVM library to accelerate their research development. Using such a library will save their time and avoid to write codes from scratch. LibSVM is one of SVM library that has been widely used by researchers to solve their problems. The library also integrated to WEKA, one of popular Data Mining tools. They also used two popular programming languages C++ and Java with three different datasets to test our analysis and experiment. The results of their research has proved that the complexity of SVM (LibSVM) is O(n^3) and the data growth will be affect and increase the time of computation.

3. PROPOSED METHOD

In this paper, a new method for object detection is presented. The main disadvantages of common methods in object detection are high time complexity, wrong object detection, not detecting objects and etc. In this paper, a new method is presented to improve some of these disadvantages. In the first step, the novel method will be designed and described. Then, it will be implemented using Matlab[1] software. The Histogram of Oriented Gradient (HOG) is used to extract features. Artificial neural network is used for classification of these features. Then the object classification is done by using combined artificial neural network with modified Adaboost classifiers (NeurAda). Here, the Adaboost algorithm needs less tuning parameters for increasing the efficiency of classification technique.

3.1 Histogram of oriented gradient

Histograms of Oriented Gradients is a shape context descriptor and can extract shape features. HOG was proposed by Dalal and Triggs [Dalal and Triggs 2005] for pedestrian detection, where they have been widely and successfully applied since, and their usage has also been extended to other applications such as eye localization [Arrspide et al. 2013]. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses

[1]http://www.mathworks.com/

overlapping local contrast normalization for improved accuracy. Recently, the HOG has been widely investigated for recognition and outperforms almost all the other feature extractors due to its robustness to illumination variation and invariance to the local geometric and photometric transformations [Tiana et al. 2016].

3.2 Artificial neural network

An artificial neural network, usually known as a neural network, is basically a mathematical model motivated by biological nervous systems like brain processes information. A neural network comprises an interconnected group of simulated neurons and it uses connectionist approach to process information for computation. Neural network works like an adaptive system, which changes its structure in learning phase. Simple and complex relationships can be easily modeled using neural networks. They are also used to find out patterns and clusters in data. An artificial neural network can be designed for a particular application, such as data classification and pattern categorization through a learning process. There are several types of neural network configurations. The arrangement of neurons to form layers and the connection pattern formed within and between layers is called the network structures [Agrawal and Agrawal 2015; **?**].

3.3 Adaboost

Boosting is used to improve the accuracy for regression and classification problems. Boosting methods are generally learned in series and these methods work best on unstable learners such as the neural network or decision tree techniques. Adaboost is one of the first practical boosting methods in which, weak learners are combined, boosted, to improve ensemble accuracy and produce a strong classifier. In classification, weak learners exhibit only a small correlation between the prediction and the true value [Kummer and Najjaran 2014].

$C(x_i) = a_1 k_1(x_i) + a_2 k_2(x_i) + \; + a_n k_n(x_i)$ (1)

$a_m = \frac{1}{2} ln(\frac{1-e_m}{e_m})$ (2)

3.4 Proposed algorithm

In this section, artificial neural network classifier will be combined with Adaboost on validation data. Object detection is consisted of three phases including training phase, validation phase and testing phase. A brief summary of each phases are given as continues:

3.4.1 *Training phase.* In this phase, 60% of all database images are taken from input as training images and HOG extracts features of these images. These features will be separately classified by artificial neural network classifier and finally, each classifier models an object detector (Figure. 1).

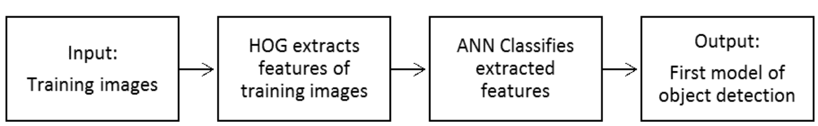

Figure 1. Training Phase

3.4.2 *Validation phase.* At the beginning of this phase, two sets of inputs are received. One of the inputs is the output model of the previous training phase and the other input is validation images which are 50% of the 40% remaining database images selected randomly. In the next step, HOG extracts features of input validation images and coefficient of Adaboost is assigned initially to zero for each validation image. Then objects in the validation images are detected by the input model. The input model detects correctly detected objects and validation phase picks up the images which are detected incorrectly and increases their by equation (4). According to new values, current object detector model will be reconsidered and corrected. If the corrected model acts better than input model enough, this phases are completed and the algorithm moves on to the next phase. Otherwise, current phase will be repeated from HOG feature extracting step again until a preferred model is gained (Figure. 2).

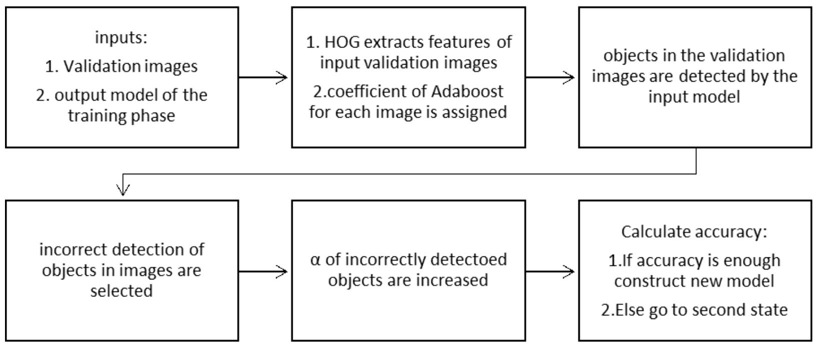

Figure 2. Validation Phase

3.4.3 *Testing phase.* In this phase (Figure. 3), there are again two inputs. First input is the 20% of the remaining database images (60% used as training data and 20% used as validation data in previous phases) and the second input is the output model of validation phase. In the next step, HOG extracts features of input testing images. Then object detection is processed with the gained model of validation phase. At the end, accuracy of the new model is calculated by equation (3).

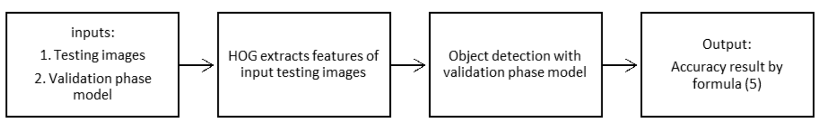

Figure 3. Testing Phase

For each object in an image, four different state might happen. If an object actually exists and the detector finds it correctly, it is a true positive state and if the detector fails to find an existing object, it is a false positive. Otherwise, true negative might happen if the detector correctly doesnt report existence of any non-existing object in an image and false negative occurs when the detector reports existence of a non-existing object. Accuracy of detector is calculated by equation below:

$$accuracy = \frac{truepositive+truenegative}{truepositive+falsepositive+falsenegative+truenegative} \quad (3)$$

3.4.4 *Output result.* Output of proposed algorithm is a model developed by training phase, then improved by validation phase and tested by testing phase. This output model detects objects more accurately than basic model of training phase and also variety of features selected by NeurAda goes up.

Table I: Running environment details

Software/Hardware	Type
CPU	Intel Core i3 M380 2.53 GH
RAM	4GB DDR3
Bus Rate	64-bit
Operation System	Windows 7 / 64-bit
VGA	ATI HD 5430
Software	Matlab R2013a

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

In this section, the proposed algorithm is implemented by Matlab software and the obtained results are compare to Pascal VOC 2011[2] which has similar demands to the proposed method of this paper. Detailed information of running environment (including software and hardware) is given in Table 1. Pascal VOC 2011 includes 28,952 different kinds of images with objects. According to importance of human beings and vehicles, 1500 images of pedestrians, 1500 images of cars and 1500 images of bicycles are taken as training/validation/testing input.

4.1 Experiments

To evaluate the proposed method, three major objects including pedestrians, cars and bicycles are expected to be detected. Image database is consisted of 1500 pedestrian/non pedestrian images, 1500 cars/non cars images and 1500 bicycle/non bicycle images. First, NeurAda is implemented to detect objects and then the obtained results are compared to the best three result of Pascal VOC 2011. For evaluation of NeurAda, 1500 pedestrian/non pedestrian images were given as input and 96% accuracy gained. Figure 4 shows accuracy rate of NeurAda in different iterations. After the fifth iteration the accuracy rate remained the same. The main reason for this behavior is that the proposed method at validation phase after a certain point concentrate more on validation data and their features, so the classifier cannot pay much attention to new input test images and behaves accurately just on data similar to validation data.

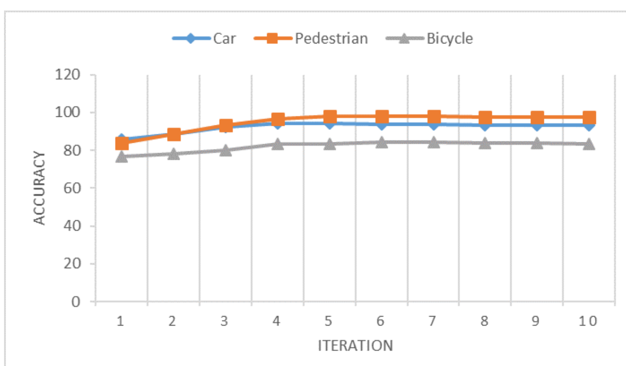

Figure 4. Object detection with NeurAda

At the end, a comprehensive comparison of new proposed method, NeurAda, and three top results of Pascal VOC 2011 is done. For each method, three category of inputs including cars, bicycles and pedestrian is tested. Each category contained 1500 images. Table 2 shows number of true positive/negative and false positive/negative which will be used in equation (3) to calculate accuracy.

[2]http://host.robots.ox.ac.uk:8080/pascal/VOC/voc2011/index.html

Table II: number of true positive/negatives and false positive/negatives in car/bicycle/human dataset

Datasets	true positive	true negatives	false positive	false negatives
Car	839	573	51	37
Bicycle	720	522	149	109
Human	707	766	15	12

Table III: Result of NeurAda implementation compared to top results of Pascal VOC 2011

Method Name	Car	Bicycle	Human
NeurAda	94.1	82.8	98.2
NLPR	85.5	82.6	93
NUSPSL	84.1	81.1	92.8
Msrauste	76.4	74.8	90.4

Table IV: Result of NeurAda implementation compared to results of Train.3 using Java (in sec.) [12]

Datasets	Java(Training Time)	Java(Testing Time)	NeurAda(Training Time)	NeurAda(Testing Time)
train.3 (400)	0.255	0.182	0.238	0.181
train.3 (800)	0.428	0.339	0.425	0.331
train.3 (1200)	0.657	0.485	0.655	0.480
train.3 (1600)	0.928	0.689	0.926	0.687
train.3 (2000)	1.557	1.185	1.552	1.180

Table 3 illustrates the final results. Among the seven compared methods, NeurAda was the most accurate in cars field with 94.1% while the best of Pascal VOC 2011 showed 85.5% accuracy which means 8.6% improvement. For bicycle detection NeurAda was the best but the improvement was not very much, only 0.2%. In pedestrian detection, NeurAda again had the highest accuracy. There was a 5.2% improvement. The main reason for this improvement is using Adaboost alongside ANN as it can be used in conjunction with many types of learning algorithms to improve their performance. The output of the ANN is combined into a weighted sum that represents the final output of the boosted classifier. Adaboost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. Adaboost is sensitive to noisy data and outliers and the final model converge to a strong learner.

In table 4, NeurAda is compared to SVM (as a famous machine learning method used for object detection). Dataset and Java running times are extracted from [12] and the same dataset ran on NeurAda. It shows a time consumption in both training and testing time in comparison to SVM.

5. CONCLUSION AND FUTURE WORK

Object detection is a very important technique in computer vision which is mainly used in many applications. There are many proposed methods to detect objects and make algorithms more accurate and several of those methods were reviewed. The main disadvantages of common methods in object detection are high time complexity, wrong object detection and not detecting objects. In this paper, a new method is presented to improve some of those disadvantages where artificial neural network was combined with Adaboost to cover weak points of previous works. At the end, new proposed methods were compared to Pascal VOC 2011 as a similar work. The recent comparison showed improvement in three field of car, bicycle and pedestrian detection. NeurAda acted as the most accurate method while detecting cars, pedestrians and bicycles. To gain further improvement, validation data could be dynamic and change per each iteration. To improve feature extracting of the propose method instead of HOG, Scale-invariant feature transform (SIFT) could be used. For detecting part base objects Deformable Part Model (DPM) can replace HOG.

REFERENCES

Abdiansah, A. and Wardoyo, R. 2015. Time complexity analysis of support vector machines (svm) in libsvm. *International Journal Computer and Application*.

Agrawal, S. and Agrawal, J. 2015. Neural network techniques for cancer prediction: A survey. *Procedia Computer Science 60,* 1, 769–774.

Arrspide, J., Salgado, L., and Camplani, M. 2013. Image-based on-road vehicle detection using cost-effective histograms of oriented gradients. *Journal of Visual Communication and Image Representation 24,* 7, 1182–1190.

Cheng, D., Wang, J., Wei, X., and Gong, Y. 2015. Training mixture of weighted svm for object detection using em algorithm. *Neurocomputing 149,* 1, 473–482.

Chitrakar, R. and Huang, C. 2014. Selection of candidate support vectors in incremental svm for network intrusion detection. *Computers and Security 45,* 1, 231–241.

Cortes, C. and V., V. 1995. Support-vector networks. *Machine Learning 20(1)*, 273–297.

Dalal, N. and Triggs, B. 2005. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 5,* 1, 886–893.

Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. 2014. Scalable object detection using deep neural networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 14,* 1, 2155–2162.

Hussain, M. A., Ansari, T. M., Gawas, P. S., and Chowdhury, N. N. 2015. Lung cancer detection using artificial neural network and fuzzy clustering. *International Journal of Advanced Research in Computer and Communication Engineering 4,* 3, 360–363.

Kong, K. K. and Hong, K. S. 2015. Design of coupled strong classifiers in adaboost framework and its application to pedestrian detection. *Pattern Recognition Letters 68,* 1, 63–69.

Kummer, N. and Najjaran, H. 2014. Adaboost. mrt: Boosting regression for multivariate estimation. *Artificial Intelligence Research 3,* 4, 64–76.

Ponti, M., Nazare, T., and Thume, G. 2015. Image quantization as a dimensionality reduction procedure in color and texture feature extraction. *Neurocomputing 15(Special)*, 1–38.

Rios-Cabrera, R. and Tuytelaars, T. 2014. Boosting masked dominant orientation templates for efficient object detection. *Computer Vision and Image Understanding 120,* 1, 103–116.

Santhi, P. and Bhaskaran, V. 2014. Detection of objects using fisher svm with modified adaboost classification technique. *Journal of Theoretical and Applied Information Technology 67,* 1, 18–26.

Tana, M., Pana, G., Wangb, Y., Zhanga, Y., and Wua, Z. 2014. L1-norm latent svm for compact features in object detection. *Neurocomputing 139(Special),* 56-64.

Tiana, S., Bhattacharyac, U., Lub, S., Sub, B., Wangd, Q., Weid, X., Lud, Y., and Tana, C. L. 2016. Multilingual scene character recognition with co-occurrence of histogram of oriented gradients. *Pattern Recognition 51,* 125-134.

Saber Shakeri received B.S. degree of Computer Science in 2012 and M.S. degree of Computer Engineering in 2015 from University of Tabriz, Tabriz, Iran. He is currently a lecturer of Faculty of Computer Engineering in Payame Noor University, Iran. His research interests are Image Processing, Computer Vision, Machine Learning and Data Mining.

Mehran Ashouraei received B.S. degree of Computer Science in 2013 form University of Tabriz, Tabriz, Iran and M.S. degree of Computer Engineering in 2015 from Islamic Azad University, Tehran, Iran. He is currently a lecturer of Faculty of Computer Engineering in Payame Noor University, Iran. His research interests are Cloud Computing, Expert Cloud, CloudSim, Genetic Algorithm, Scheduling and Image Processing.