An Optimization Method of Fusing Multiple Decisions in Object Detection

Zhu Teng(✉) and Baopeng Zhang

School of Computer and Information Technology,
Beijing Jiaotong University, Beijing, China
{zteng,bpzhang}@bjtu.edu.cn

Abstract. Object detection is widely employed in a large number of areas, such as human detection, medical image processing, etc. However, it is insufficient to use only a learning algorithm to detect objects and more techniques or models, such as a probability based approach, a part model, a segmentation model, are combined with the learning algorithm to accomplish the detection task. To this end, a fusion approach is required to balance the decisions making by multiple models. This paper proposes an optimization methodology that fuses a set of confidence outputs estimated by multiple models. Various experiments are executed and demonstrate that the proposed fusion method has a relative better performance than that of the system constituted by a single model.

Keywords: Fusing multiple decisions · Optimization method · Object detection

1 Introduction

Computer vision can be used in assorted areas, such as security (Li and Shen 2013), detection (Pedro F. Felzenszwalb et al. 2010), retrieval (Jun Wu et al. 2013), and so on, among which object detection is one of the most important areas and is widely employed in various applications. For instance, in medical image processing where the detection of anatomical objects such as the heart plays a significant role in assisting the clinicians in diagnosis, therapy planning and image-guided interventions (Yang Wang et al. 2013). The detection task has been studied by many researchers for decades and many successful results have been reported. In general, the detector that finds the bounding boxes of objects in images was realized by some learning algorithm in most works. The learning method used in object detection can be a boosting algorithm (Ivan Laptev 2009), a supported vector machine (SVM) (Scholkopf and Smola 2002), a transformation of any of them (Andreas Opelt et al. 2006) or a combination of some of them (Zheng Song et al. 2011). Besides, probability based approaches (Michael C. Burl et al. 1998) were also involved by many researchers. They are mainly utilized to encode the spatial relationship in a graphical model (Zhu Teng et al. 2014; Justin Domke et al. 2013) such as the Bayesian model (Bogdan Alexe et al. 2010), a pictorial structure (Fischler and Elschlager 1973), a tree model (Long (Leo) Zhu et al. 2010) and so on. Some other object detection methods employed a segmentation model (Bastian Leibe et al. 2008) and others build a hierarchy model to represent the object by layers

W.-C. Peng et al. (Eds.): PAKDD 2014 Workshops, LNAI 8643, pp. 29–35, 2014.
DOI: 10.1007/978-3-319-13186-3_4

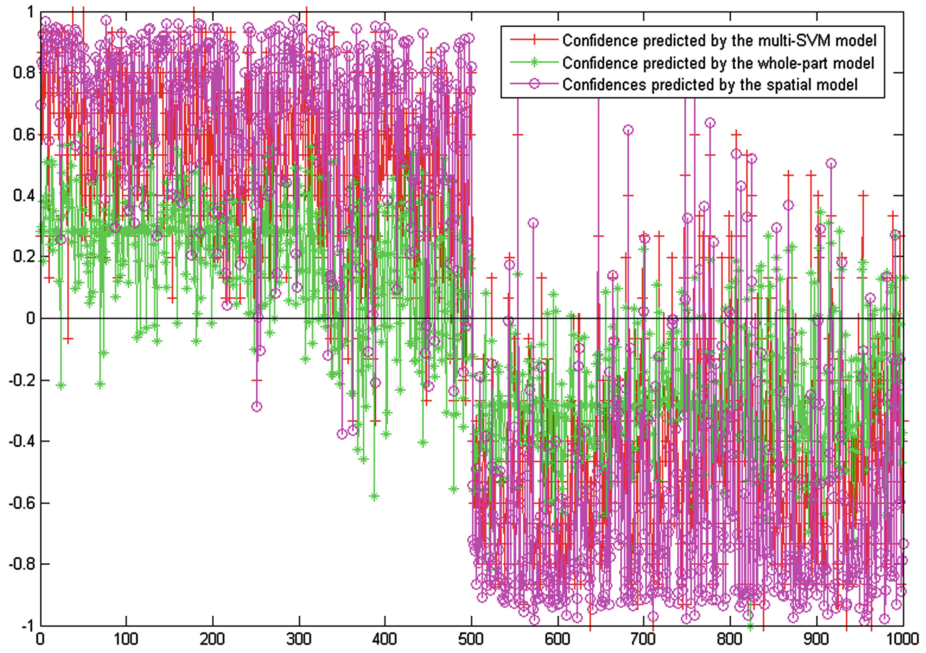

Fig. 1. An example of the confidences estimated by a learning method, a part model, and a probabiltiy based model.

(Zhu and Mumford 2006). It is much easier to detect an object if an accurate segmentation of the test image is obtained. To sum up, it is insufficient to use only a learning algorithm to detect objects and more auxiliary techniques or models (such as a probability based approach, a part model (Pedro F. Felzenszwalb et al. 2010), a graph based model (Tao Wang et al. 2012), a segmentation model, a saliency detection model (C. Lang et al. 2012), etc.) are combined with the learning method to accomplish the detection task. To this end, a fusion approach is required to balance the decisions making by multiple models.

As a motivation example, we assume three types of models, including multi-SVM model (Zhu Teng et al. 2014), part model (Pedro F. Felzenszwalb et al. 2010), spatial relationship model, are employed in the object detection task. The confidences that are estimated by these three models are based on a detection window separately. Figure 1 shows the confidences predicted by the multi-SVM model, part model, and spatial relationship model on 1000 training examples (positive: $1 \sim 500$, negative: $501 \sim 1000$) of the UIUC Image Database for Car Detection.[1] It can be seen from the figure that there are erroneous predictions for all three models, but the inaccurate confidences (such as points in the left bottom region and the top right region) predicted by these three models are not always on the same example. Therefore, if an elegant

[1] The UIUC Image Database for Car Detection is available at http://cogcomp.cs.illinois.edu/Data/Car/.

combination of these three models can be exploited, a higher accuracy of the object detection algorithm could be achieved. In brief, a fusion approach is necessary to combine these confidences in order to give a wise decision on a detection window.

In this paper, we propose an optimization approach that fuses multiple decisions made by several models to obtain higher accuracy and better performance for object detection. The remains of the paper are arranged as follows. Section 2 describes the optimization approach. Section 3 shows the experiments to demonstrate the institution of the optimization method and we come to the conclusions in Sect. 4.

2 Fusion Approach

In this section, the optimization method to combine multiple decisions is delineated. The multiple decisions are generally represented by confidences ranged from 0 to 1. If they are not, the confidences should be normalized first. Without loss of generality, three decisions making by multi-SVM model, part model and spatial relationship model are assumed and employed in this work. The goal of the fusion of the multi-SVM model, part model and spatial relationship model is to diminish the erroneous decisions making on the detection windows. An objective function is defined and an optimization method is employed to find the minimum of this objective function and the corresponding values of the variables. The optimization problem is formulated as described in Eq. (1).

$$\begin{aligned} &\underset{\alpha,\beta,\gamma,\delta}{\text{minimize}} \quad -\sum_{i=1}^{n} sign(\alpha \cdot C_{m_i} + \beta \cdot C_{p_i} + \gamma \cdot C_{c_i} + \delta) \cdot y_i \\ &\textit{subject to} \quad \alpha + \beta + \gamma + \delta = 1 \end{aligned} \tag{1}$$

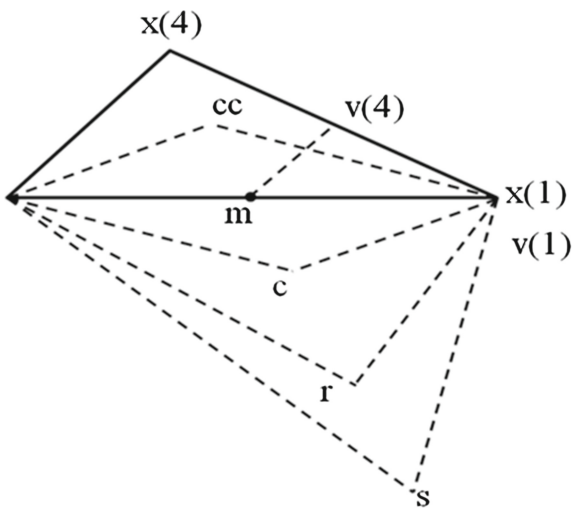

Fig. 2. Nelder-Mead (NM) simplex algorithm. The bold outline is the original simplex and the dashed outline indicates a possible new simplex.

C_{m_i}, C_{p_i}, and C_{c_i} in the objective function of Eq. (1) are the confidences on the i^{th} detection window estimated by the multi-SVM model, part model and spatial relationship model, respectively. y_i is the ground truth of the i^{th} detection window (*-1* suggests a detection window without the object and *1* indicates a detection window with the object). n is denoted as the total number of detection windows. The meaning of the objective function is the minus of the number of correctly estimated detection windows, and to minimize it is to maximize the number of detection windows that are accurately determined. α, β, γ, and δ are the optimization variables.

As the sign function and several variables are involved in the objective function, a multivariable nonlinear optimization method is required to acquire the minimum of the objective function. The method employed in this research is the Nelder-Mead (NM) simplex algorithm (J. A. Nelder, R. Mead 1965; J.C. Lagarias et al. 1998), which is an unconstrained nonlinear optimization method, and the proposed optimization problem is required to be described in an unconstrained form as shown in Eq. (2).

Table 1. NM simplex algorithm.

1) Let $x(i)$ denote the list of points in the current simplex, $i = 1,...,4$.
2) Order the points in the simplex from the lowest objective function value $f(x(1))$ to the highest $f(x(4))$. At each step in the iteration, the algorithm discards the current worst point $x(4)$, and accepts another point into the simplex. Or, in the case of step 7, it changes all *4* points with values above $f(x(1))$.
3) Generate the reflected point $r = 2m - x(4)$, where $m = \sum_{i=1}^{3} x(i)/3$, and calculate $f(r)$.
4) If $f(x(1)) \leq f(r) < f(x(3))$, accept r and terminate this iteration.
5) If $f(r) < f(x(1))$, calculate the expansion point $s = m + 2(m - x(4))$ and $f(s)$.
    a. If $f(s) < f(r)$, accept s and terminate the iteration.
    b. Otherwise, accept r and terminate the iteration.
6) If $f(r) \geq f(x(3))$, perform a contraction between m and the better one of $x(4)$ and r:
    a. If $f(r) < f(x(4))$ (i.e., r is better than $x(4)$), calculate $c = m + (r - m)/2$ and $f(c)$. If $f(c) < f(r)$, accept c and terminate the iteration. Contract outside. Otherwise, continue with Step 7).
    b. If $f(r) \geq f(x(4))$, calculate $cc = m + (x(4) - m)/2$ and $f(cc)$. If $f(cc) < f(x(n+1))$, accept cc and terminate the iteration. Contract inside. Otherwise, continue with Step 7).
7) Calculate the three points $v(i) = x(1) + (x(i) - x(1))/2$ and $f(v(i))$, $i = 2,...,4$. The simplex at the next iteration is $x(1)$, $v(2),...,v(4)$.

$$\underset{\alpha,\beta,\gamma}{\text{minimize}} \quad -\sum_{i=1}^{n} sign(\alpha \cdot C_{m_i} + \beta \cdot C_{p_i} + \gamma \cdot C_{c_i} + 1 - \alpha - \beta - \gamma) \cdot y_i \qquad (2)$$

The NM simplex algorithm belongs to the general class of the direct search method that does not utilize any derivative information. It uses a simplex of $t + 1$ points for the t-dimensional vectors $\mathbf{x}$. The algorithm first makes a simplex around the initial guess $\mathbf{x_0}$, and then updates the simplex repeatedly according to the following steps (refer to (J.C. Lagarias et al. 1998) and (J. A. Nelder, R. Mead 1965) for the details of the algorithm, and here only the necessary procedures for the optimization are given, see also Fig. 2). Since there are only three parameters in the proposed formulation, t is three in this case. The objective function is denoted by $f(x)$ ($\mathbf{x}$ denotes the variables α, β, γ) for short. Table 1 gives the details of the algorithm.

Since the NM algorithm starts at an initial estimate and finds a local solution, the NM algorithm is proceeding at different initial values a thousand times in order to avoid falling into a local optimum.

3 Experiments

In this section, the performance of the fusion method is reported on the test datasets of two categories, airplane and car, and the program is coded by Matlab. The validation dataset is distinct from both the training dataset and test dataset.

The UIUC Image Database for Car Detection contains training images, single-scale test images, and multi-scale test images, and the validation dataset is built by the single-scale test images and the test dataset is constructed by the multi-scale test images. The Caltech Airplanes dataset[2] dataset consists 1074 images and is divided into a training set (500 images), a validation set (74 images) and a test set (500 images). The three models are examined on the validation dataset, and the confidences that the multi-SVM model, part model and spatial model estimate on the images are extracted and utilized to learn optimization parameters α, β, γ. Note that there might be more than one detection window for an image from the validation dataset. The label (-1 or 1) of a detection window is following the criterion of PASCAL (Mark Everingham et al. 2010), which is obtained by comparing the detection window with the annotation (bounding box) of the corresponding image. If the overlap between the detection window and the ground-truth bounding box of the image exceeds 50 %, the detection window is considered as true. Multiple detections of the same object are considered false. For example, 4 detections of a single object (the overlap of all 4 detections is over 50 %) in an image should be counted as 1 correct detection and 3 false detections.

Table 2 presents the results on the test dataset. The performance of this experiment is evaluated by the percentage of the number of correctly detected windows to the total number of windows. A detection window is considered as true if the confidence is positive; otherwise, it is regarded as false. The only multi-SVM model of Table 2 means that only the multi-SVM confidence are used to make decisions on the examples

[2] The Caltech Airplanes dataset is available at http://www.vision.caltech.edu/html-files/archive.html.

Table 2. Accuracy comparison of the fusion method on the test dataset.

Category	Only multi-SVM model	Only part model	Only spatial model	Fusion method
Airplane	0.3155	0.4660	0.4078	0.6845
Car	0.4683	0.5000	0.6901	0.8028

in this model, and so as the only part model and the only spatial model. The confidence of the fusion method is a combination of these three confidences as discussed in Sect. 2 (α, β, γ are determined by the NM simplex algorithm). It is clear from Table 2 that the fusion method outperforms any of the other models for each category and it could further improve the performance of the object detection system.

4 Conclusions

As the accuracy of object detection is demanded higher and higher, detection by only a learning algorithm is insufficient and multiple models are entailed to be combined. In this paper, we propose an optimization method to combine multiple decisions making by a learning method and some other models or techniques. The experiments on the benchmark datasets demonstrate that the fusion method performs better than any single model that composes the fusion approach.

Acknowledgements. This work was supported by the Fundamental Research Funds for the Central Universities with grant number 2014JBM040 and Natural Science Foundation of China (61370070).

References

Opelt, A., Pinz, A., Zisserman, A.: Incremental learning of object detectors using a visual shape alphabet. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 3–10 (2006)

Leibe, B., Leonardis, A., Schiele, B.: Robust Object Detection with Interleaved Categorization and Segmentation. Int J Comput Vis **77**, 259–289 (2008). doi:10.1007/s11263-007-0095-3

Scholkopf, B., Smola, A.J.: Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning). The MIT Press, Cambridge (2002)

Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2010)

Lang, C., Liu, G., Yu, J., Yan, S.: Saliency detection by multitask sparsity pursuit. IEEE Trans. Image Process. **21**(3), 1327–1338 (2012)

Laptev, I.: Improving object detection with boosted histograms. Image Vis. Comput. **27**, 535–544 (2009)

Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**, 308–313 (1965). doi:10.1093/comjnl/7.4.308

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. **9**(1), 112–147 (1998)

Wu, J., Shen, H., Li, Y.-D., Xiao, Z.-B., Ming-Yu, L., Wang, C.-L.: Learning a hybrid similarity measure for image retrieval. Pattern Recogn. **46**(11), 2927–2939 (2013)

Domke, J.: Learning graphical model parameters with approximate marginal inference. PAMI, 35 (10), pp. 2454–2467 (2013) (to appear)

Zhu, L., Chen, Y., Yuille, A., Freeman, W.: Latent hierarchical structural learning for object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2010)

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**, 303–338 (2010). doi:10.1007/s11263-009-0275-4

Fischler, M.A., Elschlager, R.A.: The representation and matching of pictorial structures. IEEE Trans. Comput. **c-22**(1), 67–92 (1973)

Burl, M.C., Weber, M., Perona, P.: A probabilistic approach to object recognition using local photometry and global geometry. In: Proceedings of European Conference on Computer Vision, pp. 628–641 (1998)

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. 32(9), 1627–1645 (2010)

Zhu, S.-C., Mumford, D.: A stochastic grammar of images. Found. Trends Comput. Graph. Vis. **2**(4), 259–362 (2006). doi:10.1561/0600000018

Wang, T., Dai, G., Ni, B., Xu, D., Siewe, F.: A distance measure between labeled combinatorial maps. Comput. Vis. Image Underst. **116**(2012), 1168–1177 (2012)

Wang, Y., Georgescu, B., Chen, T., Wen, W., Wang, P., Xiaoguang, L., Lonasec, R., Zheng, Y., Comaniciu, D.: Learning-based detection and tracking in medical imaging: a probabilistic approach. Lect. Notes Comput. Vis. Biomech. **7**, 209–235 (2013)

Li, Y., Shen, H.: On identity disclosure control for hypergraph-based data publishing. IEEE Trans. Inf. Forensics Secur. **8**(8), 1384–1396 (2013)

Song, Z., Chen, Q., Huang, Z., Hua, Y., Yan, S.: Contextualizing object detection and classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011)

Teng, Z., Zhang, B., Kim, O., Kang, D.-J.: Regional SVM classifiers with a spatial model for object detection. In: International Conference on Computer Vision Theory and Applications, Lisbon, Portugal (2014)