An Object Detection Algorithm based on Deformable Part Models with Bing Features

Chunwei Li[1, a] and Youjun Bu[1, b]

[1] National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China;

[a]lichunwei15@126.com, [b]13017681302@163.com

Keywords: Object detection, deformable part models, binarized normed gradients feature.

Abstract. To solve the problem that the positioning strategy with sliding window approaches requires exhaustive search in feature pyramids, the paper proposes an object detection algorithm based on deformable part models with Bing features to help object detection. First of all, input images are preprocessed with the objectness detection algorithm with Bing features and a set of potential windows that may contain target objects are obtained, and then the deformable part model is regarded as the class-specific detector to match potential windows, at last Non-Maximum Suppression is used to merge and reduce window areas of results to obtain final detection results. The experimental results on Pascal VOC 2007 dataset show that the algorithm in the paper outperforms the original DPM in 19 out of 20 classes, achieving an improvement of 2.7% mAP.

Introduction

Object detection aims to determine whether there is a class of objects in static image we are interested in and if so, given the information about its size, location etc. and is one of the key problems in the field of image processing and computer vision. Because of the existence of complex background, occlusion, illumination and inter-class differences, object detection is a challenging problem.

After decades of development, especially after the Deformable Part Model (DPM) is put forward, the technology of object detection has made great progress and there are a large number of algorithms proposed to improve DPM, including the development of low-level features: sparse codes[1] and Convolutional Neural Networks[2]; the development of training processes of samples: strongly-supervised learning[3] and semi-supervised learning[4]; the development of sample annotations: based on 3D geometry information[5] and so on. But above methods are still dependent on the sliding window search strategy to locate objects (Fig. 1). Although this method guarantees the recall rate, real objects appear in limited positions and most detected windows are invalid for visual object detection and so many windows would increase the false-positive rate .

Some scholars aimed to further improve the detection effect by means of improvement of sliding window method. [6] combined SLIC super-pixels with sliding window detectors to split the image into foreground and background channels. Others combined object detection algorithms with objectness detection algorithms so as to guide the positioning process to improve the detection results. [8] integrated a variety of visual cues in a Bayesian framework to detect proposals. [9] proposed to utilize diverse sampling methods to guide the search process. [10] took use of edge information to determine proposals.

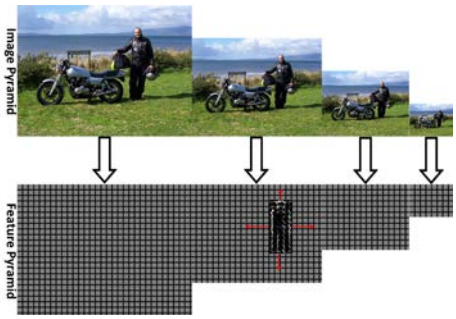

Fig. 1 The example of sliding window approaches

But these methods are insufficient realization of complex and large amount of computation, for this reason we put forward a proposal positioning method based on Bing features and apply it to deformable part models, namely Improved Deformable Part Models with Bing features (BDPM). Compared with previous work, our work has following three aspects: (1) utilizing the proposal positioning method based on Bing features to help detection results; (2) proposing a screening method of proposals based on the template size and the objectness score; (3) the algorithm in the paper outperforms the original DPM in 19 out of 20 classes on Pascal VOC 2007 dataset, achieving an improvement of 2.7% mAP (mean average precision).

Positioning Proposals Based on Bing Features

NG Features. According to [9], for the object with a well-defined closed boundary, at the normal gradient space, after scaling the corresponding image window to a small fixed size, there is a strong correlation between the gradient magnitude and then Support Vector Machine (SVM) classifier can be able to distinguish objects and background. This characteristic can be used to improve the recall rate of object detection. In order to detect proposals, after scaling corresponding images to several predefined different sizes and computing the normed gradients of every resized image, the values in an 8×8 region of these resized normed gradients maps are defined as the normed gradients feature(NG feature).

Positioning Proposals. We obtain the location information of a series of sampling windows of different sizes and corresponding objectness scores through the following operations:

(1) NG features are extracted from corresponding windows of real objects and random background windows as positive samples and negative samples respectively; and then linear SVM is used to train to obtain the corresponding SVM classifier vector $\mathbf{w} \in \mathbb{R}^{64}$;

(2) We scan over several predefined window sizes of input images to sample windows that may contain objects which is defined as:

$$f = (i, x, y) \tag{1}$$

where f, i and (x, y) respectively indicate the location, size and position of a window;

(3) The above sampled windows f were normalized to 8×8 and is scored with the linear classifier vector $\mathbf{w}$:

$$s_f = \langle \mathbf{w}, g_f \rangle \tag{2}$$

where g_f indicates NG feature of the window and s_f indicates the classifier output score;

(4) We use NMS (Non-maximum suppression) to select a small set of windows for each size i. Then linear SVM is used to train the coefficient v_i and bias term t_i for each size i using the computed score s_f as samples and then the objectness score is defined as:

$$O_f = v_i \cdot s_f + t_i \tag{3}$$

Because some sizes (*i.e.* 256×32) are less likely than others (*i.e.* 128×128) to obtain objects.

In practical applications, in order to accelerate the speed of NG features, we use binary approximation to binarize NG feature defined as Bing features. For the linear model $\mathbf{w} \in \mathbb{R}^{64}$, we use a set of basis vectors to approximate $\mathbf{w}$ as:

$$\mathbf{w} \approx \sum_{j=1}^{N_w} \beta_j a_j \tag{4}$$

where N_w indicates the number of basis vectors, $a_j \in \{-1,1\}^{64}$ indicates the basis vector and $\beta_j \in \mathbb{R}$ indicates the corresponding coefficient. In order to facilitate the processing, setting $a_j^+ \in \{0,1\}^{64}$, and then the basis vector is defined as $a_j = a_j^+ - \overline{a_j^+}$. In order to save storage space, the normed gradient values (stored in a byte value) is approximated by using the top N_g binary bits of the byte:

$$g_f = \sum_{k=1}^{N_g} 2^{8-k} b_{k,f} \tag{5}$$

where $b_{k,f}$ indicates corresponding values of the top N_g binary bits (*i.e.* the normed gradient value is 180 and can be represented as a binary number 10110100. When the value of N_g is 4, then the normed gradient value is approximated as $180 \approx \sum_{k=1}^{N_g} 2^{8-k} b_{k,f} = 2^{8-1} \times 1 + 2^{8-2} \times 0 + 2^{8-3} \times 1 + 2^{8-4} \times 1$ by using the top 4 binary bits 1011.). In Fig. 2, we use different colors to distinguish objectness scores of proposals.

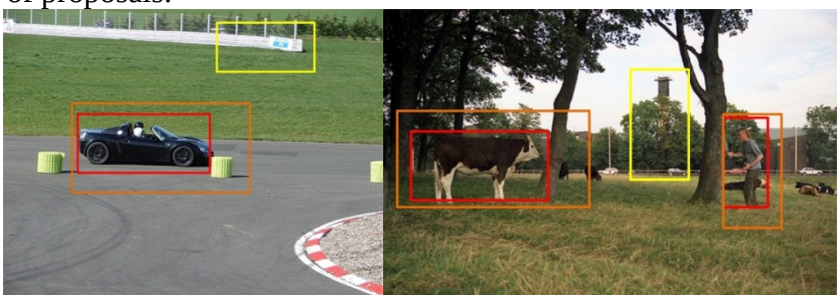

Fig.2 The detection results of objectness detetcion (The red boxes indicate the results of highest scores, the orange ones indicate the results of middle scores and the yellow ones indicate the results of lowest scores.)

Improved Deformable Part Models with Bing Features

Overview of Deformable Part Models. Deformable Part Models is mainly consisted of three parts: (1) a rough root filter ω_0 covering the whole object to describe the contour; (2) a number of high resolution filter filters $\omega_i (i = 1, \dots, n)$ used to describe the detail features; (3) the deformation model $d_i (i = 1, \dots, n)$ indicates the position of the part relative to the detection window and the corresponding deformation cost.

An object hypothesis is defined as $\{p_0, p_1, \dots, p_n\}$ where p_0 indicates the position of w_0 and $p_t (t = 1, \dots, n)$ indicates the position of the t-th part. Therefore the overall score of the object hypothesis is equal to the sum of the scores of each filter minus corresponding deformation costs:

$$\text{score}_f(p_0, \dots, p_n) = \omega_0^T \phi_\alpha(p_0, H_f) + \sum_{i=1}^{n} \omega_i^T \phi_\alpha(p_i, H_f) - d_i^T \phi_d(p_i, p_0) + b \tag{9}$$

Where H_f indicates the feature pyramid of the proposal. ϕ_α and ϕ_d respectively indicates the HOG feature and the deformation feature between part and root filters. b is a bias to distinguish different components. We detect the position of the target object according to the most likely part configuration:

$$\text{score}_f(p_0) = \max_{p_1, \dots, p_n} \text{score}_f(p_0, \dots, p_n) \tag{10}$$

Algorithm Method. As stated earlier, original DPM computes the overall score at each window position of each scale in order to detect interested objects. Therefore we propose Improved Deformable Part Models with Bing features to help object detection. Proposals and corresponding scores $(f_1, \dots, f_n)$ are obtained by the objectness detection algorithm based on Bing features and then a set of reliable object windows $(\widehat{f_1}, \dots, \widehat{f_m})$ are screened out from the proposals based on the size and position of the original root filter and then are resized to the approximate scales between the maximum and minimum size of the template instead of generating the complete feature pyramid. DPM is used as the class-specific detector to determine whether there is objects of corresponding classes and at last Non-Maximum Suppression is used to merge and reduce the window areas of the

results to obtain final detection results to avoid multiple overlapping windows. The diagram of BDPM is shown in Fig. 3.

In the concrete implementation, we propose a screening method for proposals based on the size of the template and the objectness score to further prune proposals with small possibilities. On the one hand the real object instances appear in different sizes and scales in actual scenes and so we can regard the aspect ratio of the proposal as one of basic properties of objects. If the deviation degree is too big, we think that the proposal doesn't contain a generic object or there is a large scale gap. On the other hand we expect the proposals which are false positives have a low objectness score O_f. Thus in the process of detection, we regard the objectness score as a reference value. These proposals with higher O_f should be paid to more attention.

Based on above intuition, we linearly combine the class-specific score $\text{score}_f(p_0)$ of a proposal with its objectness score O_f: $\text{score}_f^* = \text{score}_f(p_0) + \alpha \cdot O_f$. This combination can be used instead of the class-specific score where the weight α controls the importance of the objectness score.

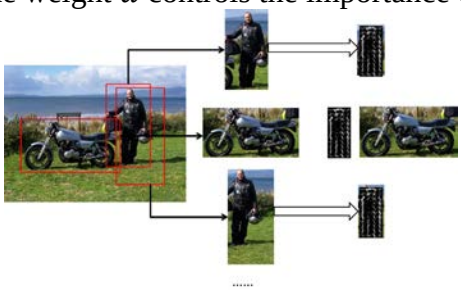

Fig.3 Improved Deformable Part Models with Bing Features(The first line indicates the successful detection result, the second line indicates the proposal of huge difference compared with the size of the model and the third line indicates unsuccessful detection result.)

Experimental Results and Analysis

In order to evaluate the performance of the proposed algorithm, the average precision (AP) is used as the evaluation index. Experiments are conducted on PASCAL VOC 2007 dataset which is a generic object detection dataset which includes 20 categories and 4952 test images. There are a total of 14976 manual annotation objects in the dataset.

Effects of the Number of Proposals. From the previous description the number of proposals and the threshold of the deviation degree of the aspect ratio of the proposal relative to that of the template used during selecting proposals are two important parameters. We first discuss the effect of the threshold of the deviation degree rat on detection results. rat is defined as:

$$\text{rat} = \left|\text{obj}_f - \text{mod}\right|/\text{mod} \tag{11}$$

Where obj_f and mod respectively indicates the aspect ratio of the proposal and that of the template. According to [9], using 1000 proposals basically already contains most of generic objects in the image, thus we first discuss the effect of different threshold of rat on detection result when the number of proposals is 1000. The experimental results of two classes are shown in Fig. 4. It can be found that before the threshold reaching 0.7, the AP value increases gradually and after that the AP value tends to be stable. Thus the threshold of rat is set to 0.7 in later experiments.

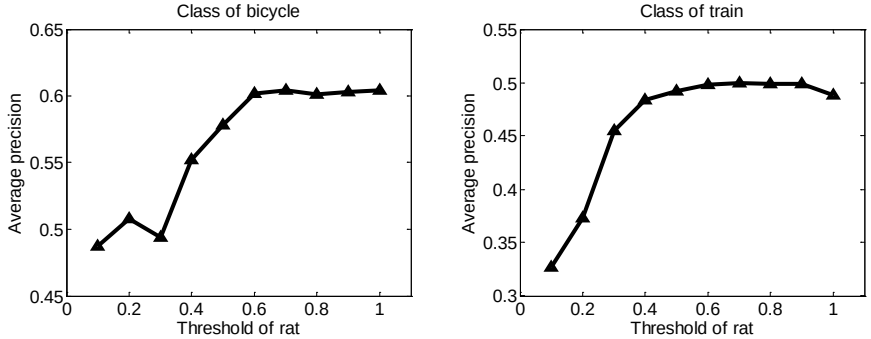

Fig. 4 Effects of the threshold of rat on detection results

After the threshold of rat is set, we use different number of proposals to describe the specific effect on detection result. The experimental results of two classes are shown in Fig 5. We can find that when the number is increased from 50 to 200, the AP value increases steadily but the tendency decreases gradually. When the number is 200, the AP value tends to be stable or even degrades slightly. This is mainly due to with the increase of the number of proposals although proposals may contain new generic objects or more precise target ranges, the possibility is gradually reduced and the contribution to final detection results is reduced. At the same time, it may increase the probability of false positives. Therefore in later experiments the number is set to 200.

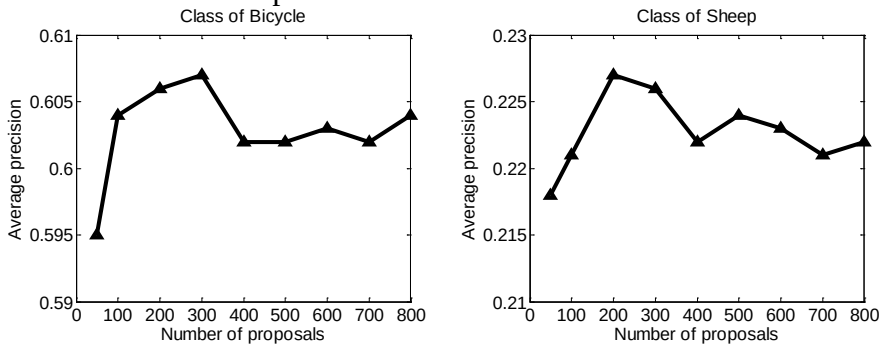

Fig. 5 Effects of the number of object windows on detection results

Effects of the weight α. On the basis of the above, we discuss the effect of the value of α for detection results and the experimental results are shown in Fig. 6. It can be found that when the value of α is set to 0.2, the AP value has achieved ideal result. Continuing to increase α even has side effects on detection results.

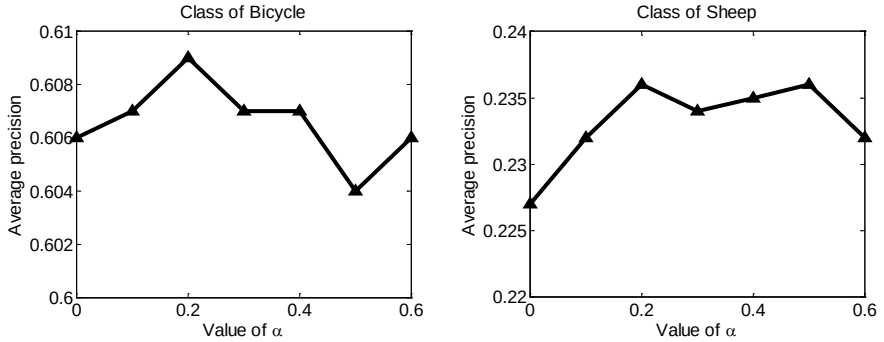

Fig. 6 Effects of the value of α on detection results

Experimental Results of Pascal VOC 2007. In Tab. 1, we compare proposed BDPM with original DPM[7] and several other objectness algorithms including DPM with Bounding Box (BB), MDPM[8] and SegDPM[6]. The sixth row indicates the experimental results that combine the class-specific score with its objectness score expressed by BDPM_{α}.We are able to see that the detection results of BDPM have outperformed the original DPM model in most classes and there are two main reasons.

Table 1 Results of different algorithms on PASCAL2007 dataset

Class	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow
DPM	33.0%	54.7%	11.0%	15.0%	23.7%	44.4%	41.6%	26.5%	16.1%	27.3%
BB	32.1%	59.9%	10.6%	13.5%	24.9%	47.7%	48.9%	27.5%	17.4%	27.0%
MDPM	32.1%	59.3%	11.1%	13.2%	**25.1%**	**49.8%**	49.0%	29.2%	17.4%	27.7%
SegDPM	33.8%	**61.0%**	**14.7%**	14.2%	23.6%	45.6%	49.1%	30.1%	18.2%	**29.2%**
BDPM	36.4%	60.6%	12.9%	15.5%	24.4%	48.5%	50.8%	30.9%	19.4%	28.1%
BDPM_{α}	**36.9%**	60.9%	12.8%	**15.7%**	24.9%	49.6%	**51.6%**	**30.9%**	**19.8%**	28.7%

Diningtable	Dog	Horse	Motorbike	Person	Plant	Sheep	Sofa	Train	Tv	Mean
29.4%	13.8%	56.4%	49.3%	36.3%	14.1%	18.6%	34.7%	45.6%	39.2%	31.5%
31.1%	14.4%	57.4%	50.0%	38.4%	13.3%	20.1%	36.8%	47.5%	42.3%	33.0%
32.7%	17.1%	59.0%	51.3%	38.8%	**15.9%**	16.9%	37.2%	48.2%	43.2%	33.7%
31.8%	17.3%	**59.8%**	52.3%	37.9%	14.4%	22.6%	**41.2%**	**50.7%**	42.9%	34.5%

35.0%	**17.6%**	59.1%	52.3%	41.8%	14.2%	22.7%	38.2%	50.1%	44.2%	35.1%
35.7%	17.5%	59.7%	**52.9%**	**42.6%**	14.6%	**23.6%**	39.1%	50.6%	**44.8%**	35.7%

Compared with the positioning strategy with sliding window approaches, BDPM conducts detection in a small set of potential windows and it helps to decrease the number of windows need to detect by several orders of magnitudes. These windows meanwhile have already contained most of target objects, therefore on the basis of ensuring the recall rate, a substantial reduction of the scope and the number of windows decreases the probability of false positive. In the case of same recall rate, the improvement of precision rate leads to the increasing of AP value.

Compared with original DPM, BDPM believes that generic objects tend to appear in the central position of potential windows and the possibility appears in the edge position of windows is smaller. Specific to each potential window, it would be resized to the approximate scales between the maximum and minimum size of the template instead of detecting in the complete feature pyramid which further decreases the probability of false positive.

Conclusions

We present an object detection algorithm based on Improved Deformable Part Models with Bing Features to help object detection. The future work is to further utilize the visual cues in potential object windows and improve the accuracy the proposal containing the generic object. It will be helpful to design better screening algorithm of proposals in order to provide a better basis for class-specific detection. Meanwhile the proposal positioning method also can be combined with other object detection algorithms such as R-CNN[11] to further improve the efficiency and accuracy of object detection.

Acknowledgments

This work is supported by 2014BAH30B01 of the Chinese National Key Technology Support Program and 61521003,61572519 of the Chinese National Natural Science Foundation.

References

[1] Ren Xiao-feng and Ramanan D. Histograms of Sparse Codes for Object Detection[C], Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, 2013: 3246-3253.

[2] Savalle P, Tsogkas S, Papandreou G, et al. Deformable Part Models with CNN Features [C], Proceedings of the 13th European Conference on Computer Vision Parts and Attributes Workshop, Zurich, 2014: 1-4.

[3] Azizpour H and Laptev I. Object Detection Using Strongly-Supervised Deformable Part Models [C], Proceedings of the 12th European Conference on Computer Vision, Firenze, 2012: 836-849.

[4] Branson S, Perona P, and Belongie S. Strong Supervision From Weak Annotation: Interactive Training of Deformable Part Models [C], Proceedings of the 13th International Conference on Computer Vision, Barcelona, 2011: 1832-1839.

[5] Pepik B, Stark M, Gehler P, et al. Multi-view and 3D Deformable Part Models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015,37(11) : 1-14.

[6] Trulls E, Tsogkas S, Kokkinos I, et al. Segmentation-aware Deformable Part Models [C], Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbs, 2014: 4321-4328.

[7] Felzenszwalb P, Girshick R, McAllester D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,

2010 ,32 (9): 1627-1645.

[8] Bogdan Alexe, Thomas Deselaers, Vittorio Ferrari. Measuring the Objectness of Image Windows[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11):2189-2202.

[9] Cheng Ming-ming, Zhang Zi-ming, Lin Wen-yan, et al. BING: Binarized Normed Gradients for Objectness Estimation at 300fps [C], Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbs, 2014: 3286-3293.

[10] Lawrence Zitnick, and Piotr Dollár. Edge Boxes: Locating Object Proposals from Edges[C]. Proceedings of the 13th European Conference on Computer Vision, Zurich, 2014: 391-405.

[11] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C], Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbs, 2014: 580 - 587.