A New Trademark Detection Method via Trademark Confidence Score of MSERs

Yang Zheng[a], Jie Liu*[a], Yuan Zhang[a], Shuwu Zhang[a,b], Qing Li[c]

[a]Institute of Automation, Chinese Academy of Sciences, Beijing, China; [b]AICFVE, Beijing Film Academy, Beijing, China; [c]School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China

ABSTRACT

This paper proposes a new algorithm to provide high quality potential trademark locations for trademark detection. In real-world circumstances, trademark regions often possess some distinctive, invariant and stable properties which can be gained effectively and efficiently by Maximally Stable Extremal Regions (MSERs). Based on this observation, we design Trademark Confidence Score (TCS) for adaptive MSERs in the images. Then a window refinement algorithm is proposed to retain the high-quality candidate windows generated by Selective Search (SS). Experiments on FlickerLogos-27 and our own dataset demonstrate that our algorithm can significantly reduce the number of candidate proposals produced by SS with little sacrifice of recall for trademarks. Moreover, for trademark detection, our algorithm has better performance while reducing the computational cost of detection.

Keywords: trademark detection, Maximally Stable Extremal Regions, Trademark Confidence Score, Selective Search

1. INTRODUCTION

Trademarks, a symbol of identification for companies, products and organizations, can be regarded as objects with a planar surface and are significantly valuable in many practical scenarios of modern marketing, advertising and trademark registration. The detection of trademark in real-world images is a meaningful sub-topic in object detection, due to its huge potential market in both commercial and civil application [1]. However, the appearance of trademarks can scale up to thousands of classes, and different trademarks may exhibit totally different morphological characteristics, which bring tremendous challenge to the existing detection methods [2].

A few deep methods [3, 4, 5] have been recently proposed by exploiting the state-of-the-art object detection models. Recently, Uijlings et al. [6] proposed selective search (SS) based on image segmentation, which is class-independent and achieves promising results in providing candidate proposals for generic object detection. However, there are still 362 candidate windows even with “single strategy”. It demands massive time to make further judgement on all these windows. Bianco [7] exploit selective search algorithm to provide a set of regions likely to contain an instance of the object of interest, i.e. logos in our case. The proposed regions will be disambiguated by the neural network that comes afterward. Maximally Stable Extremal Regions (MSERs) [8] is often used to extract stable areas as candidate characters in text detection [9, 10, 11]. Similar to characters, trademarks also stand in stark contrast to the background in natural scenes. Therefore, MSERs are valuable prior knowledge for detecting trademark regions. However, the trademark has its own unique properties. As shown in Figure 1, trademarks have tens of thousands of classes, which may consist of only an area (e.g. McDonalds) or multiple regions (e.g. Starbucks) and may have dense distribution (e.g. Vodafone, Tencent) or sparse distribution (e.g. Adidas, Baidu). The trademark regions are highly different. As a result, there are no universal rules to aggregate MSERs of trademarks. It is almost impossible to take each MSER separately for trademark detection. To sum up, SS requires location filtration to speed up the detection process. When applying MSERs for trademark detection, we need to aggregate different parts of trademarks to obtain possible locations.

To achieve high quality candidate proposals for trademark detection, in this paper we integrate the merits of selective search and MSERs for generating trademark proposals. The rest of this paper is arranged as follows: In Section 2 we present the details of our work, and experimental comparisons are shown in Section 3. Finally, we give the conclusion and further work in Section 4.

*jie.liu@ia.ac.cn

Figure 1. Examples of trademarks: McDonalds, Pepsi, Vodafone, Tencent, Starbucks, Adidas and Baidu.

2. PROPOSED ALGORITHM

2.1 Adaptive MSERs Detection

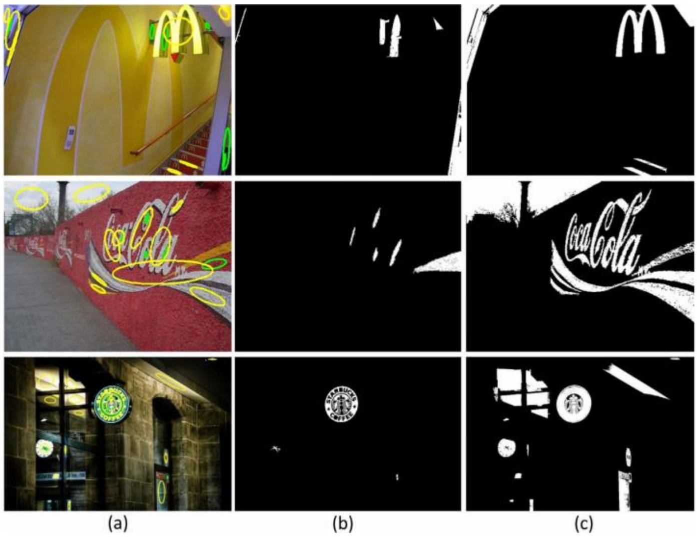

Figure 2. (a) Adaptive MSER detection results of real-world images. The dark on bright MSERs are marked by green ellipses and the bright on dark MSERs are marked by yellow ellipses. (b) Bright on dark MSERs by adaptive MSERs detection. MSERs are white and other regions are black. (c) Bright on dark MSERs by original MSERs detection. MSERs marked in the red rectangles in (c) are removed.

Trademarks are made of characters, graphics, letters, numbers, three-dimensional signs and other connected components. When they appear in natural images, trademark regions stand in stark contrast to their surroundings. They can be regarded as distinguished regions (DRs). When we observe an image, it is easy to be attracted by these DRs. Therefore, they will provide important clues for trademark detection. MSERs [8] is one of best region detectors, which can be used to extract the DRs in the images. In text detection, many researchers employ MSERs as candidate regions of text characters [9, 10, 11]. The scenes where texts appear are similar to trademarks, so trademarks can be seen as a special kind of "texts". Therefore, MSERs can also be used to offer key information.

The MSERs extraction is originally controlled by a predefined threshold Δ, which decides how stable the image regions are. The higher Δ is, the fewer MSERs are. It is not appropriate to set a fixed Δ for extracting trademark MSERs because of the variety of the trademark appearance. For "simple" images, Δ should be lower to avoid eliminating the trademark MSERs. For "complex" images, there are many extremal regions, but most of which may not be trademark regions. In this case, Δ should be larger so as to remove the superfluous non-trademark regions.

Since SIFT key points can reflect the image "complexity" to some extent [12, 13], we design adaptive Δ for images to gain optimum MSERs as:

$$\Delta=\mu+\lambda*\frac{KN}{Area} \tag{1}$$

where μ is a constant, which denotes stable part of Δ. *KN* is the number of SIFT key points in the image, and Area is the product of height and width of the image. λ is a parameter to control the weight of adaptive part in Δ. When $\lambda = 0$, Δ becomes traditional fixed case. Through cross validation, we find that $\mu = 10$ and $\lambda = 2700$ will lead to the best performance of the proposed algorithm. Figure 2 illustrates adaptive MSERs and original MSERs detection results.

2.2 Trademark Confidence Score for MSERs

In this section, we introduce a new concept named trademark confidence score (TCS) to indicate the confidence of the extracted MSERs as trademark regions. In this paper, the TCS for each MSER is given by a classifier. Suppose $X = \{x_1, x_2, ..., x_n\}$ retains *n* training samples, and $Y = \{y_1, y_2, ..., y_n\}$ is the corresponding label set. The positive training samples are image patches which only contain trademark logos, and the negative samples are images patches without trademark regions. Wherein, $x_i \in R^d$ denotes the HOG feature of the *i*-th sample, and $y_i \in \{0,1\}$ is the label of x_i . The loss function of the classifier is defined as：

$$\min_{w} \frac{1}{n}\sum_{i=1}^{n}\log(1+e^{-y_i w^T x_i})+\gamma\|w\|_2^2 \tag{2}$$

where *w* denotes classification parameters, and γ is the coefficient to control regularization term. The above Logistic regression model can be solved by stochastic gradient descent (SGD). When the classification parameters are obtained, we can calculate a trademark confidence score for each MSER as:

$$TCS(\mathrm{x}^*) = \frac{1}{1+e^{-w^T x^*}} \tag{3}$$

where x^* is the HOG feature of an image patch generated by a MSER. Generally, it is reasonable to deem that a MSER with low score may not be a trademark

2.3 The Window Refinement Algorithm Based on TCS

For trademarks with one region such as “McDonald” (Figure 2), the whole trademark can be extracted by a single MSER. However, there also exist some trademarks with many separated stable regions such as “Coca-Cola” (Figure 2), in which a single MSER only contain parts of the trademark. If we only use one MSER to select windows, some parts of the trademark will retain while the whole trademark may be discarded. Therefore, it is unreasonable to handle various scenarios by a single strategy.

After calculating the TCS for MSERs, we can use them to evaluate the windows produced by selective search. In real-world images, each trademark may contain one or several MSERs. Similarly, each SS window may intersect with one or more MSERs. Since both SS and MSERs windows are used to generate possible trademark locations, the intersection-over-union (IOU) ratios between them can indicate the importance of the proposals to same extent. Thus, we combine IOU and TCS to refine windows generated by SS as shown in Algorithm 1.

Algorithm 1 The window refinement based on IOU and TCS of MSERs

```
Require:
    m windows generated by selective search, n bounding boxes of MSERs, the corresponding TCS
    vector $S_{TCS} \in R^n$, and predefined threshold $\tau$ for pruning the windows.
Ensure:
   potential trademark proposals.
   1: Calculate IOUs between SS windows and adaptive MSERs bounding box to get $mat_{IOU} \in R^{m \times n}$ .
   2: Calculate WCS for SS windows as
          $S_{WCS} = Mat_{IOU} \bullet S_{TCS}$          (4)
   3: Prune the windows with WCS smaller than $\tau$ .
   4: return the remained windows as potential trademark proposals.
```

$mat_{IOU} \in R^{m \times n}$ denotes the IOU matrix between *m* SS windows and *n* adaptive MSERs bounding boxes in an image. The entry mat_{IOU} is IOU between the *i*-th SS windows and the *j*-th MSERs. After obtaining mat_{IOU}, we calculate window confidence score (*WCS*) for windows as (4). $S_{wcs} \in R^{m}$ can be deemed as the confidence vector for SS windows as a trademark. A window with higher *WCS* tends to be a trademark. Thus, by giving a certain threshold τ, the windows with *WCS* smaller than τ are discarded.

3. EXPERIMENTS

3.1 Data and Evaluation Metric

We evaluate our proposed algorithm on FilckerLogos-27 dataset [2] and our own dataset against the alternative proposals generating algorithms: SS, Edge box [14] and BING [15]. FilckerLogos-27 contains 27 classes of trademarks with total 1079 real-world images. They are: Adidas, Apple, BMW, Citroen, Coca-Cola, DHL, FedEx, Ferrari, Ford, Google, Heineken, HP, McDonalds, Mini, NBC, Nike, Intel, Pepsi, Porsche, Puma, Red bull, Sprite, Starbucks, Texaco, UNICEF, Vodafone and Yahoo. There are 4531 trademark samples in the 1079 images. We use the training set and testing set introduced in [2]. We randomly select 3000 positive samples and 5000 negative samples (drawn from the background) as training set. The images with the remaining 1531 samples are testing set. We also collect real-world images with trademarks to form our own dataset for further evaluation. The dataset contains 9 trademarks with total 490 images including: Apple, HSBC, ABC, SDP Bank, TAIPING Life, Sina, China Merchants Bank, China Telecom and China Unicom. Each image contains only one trademark. We randomly select trademarks from 300 images and 500 background patches for training, and the 190 images are used for evaluation.

In order to evaluate the accuracy of possible trademark locations properly, we adopt the criteria in [6]: Average Best Overlap (ABO), Mean Average Best Overlap (MABO) scores and the number of candidate windows. For a specific class *c*, ABO is calculated as

$$ABO = \frac{1}{|G^c|} \sum_{g_i^c \in G^c} \max_{l_j \in \Omega} Overlap(g_i^c, l_j) \tag{5}$$

where G^c is the ground truth set of class *c*, and the element g_i^c in G^c denotes the *i*-th ground truth bounding box. $|G^c|$ returns the number of ground truth for class *c*. Ω retains the set of object proposals, and l_j denotes the object hypotheses windows. The $Overlap(.,.)$ function measures the intersection between the object hypotheses and the ground truth bounding box in the corresponding image. Then the MABO can be calculated as the mean ABO over all the classes in the given dataset.

3.2 Parameter Setting

To calculate TCS for MSER, each input sample is resized to 32 × 32. The regularization coefficient γ in (2) is empirically set as 5. Additionally, the threshold τ in Algorithm 1 plays an important role in controlling the tradeoff between MABO and the number of proposals. The smaller τ may lead to higher MABO, but cannot prune much meaningless proposals. On the other hand, when τ is set bigger, some promising proposals may be removed. Therefore, we design a sensitivity analysis to select τ.

The *WCS* in Algorithm 1 can be viewed the confidence score for each proposal generated by SS. We use a discrete set $\Lambda = \{0.1\mu, 0.2\mu, ..., 0.9\mu\}$ to parameterize τ, where μ denotes the mean of S_{WCS} in Algorithm 1. The trend of MABO and the proposal number under different values of τ can be seen in Figure 3.

A larger τ results in the smaller number of proposals. However, the MABO is getting worse at the same time as shown in Figure 3. Figure 3 shows that $\tau = 0.2\mu$ is an inflection point of the curves. The MABO decreases more rapid when $\tau > 0.2\mu$. Although there are other similar inflection points when $\tau > 0.2\mu$, we prefer a relatively high MABO for further detection. Thus, taking the tradeoff into consideration, τ is adaptively set as 0.2μ.

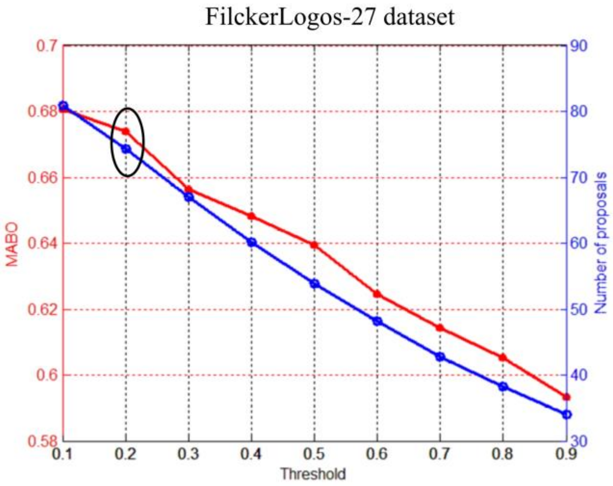

Figure 3. Sensitivity analysis about the τ on FlickerLogos-27. To balance the performance of MABO (left y-axis, red line) and number of proposals (right y-axis, blue line), τ is adaptively set as 0.2μ.

3.3 Quantitative Comparison of Proposals

Table 1 shows the MABO and number of proposals given by different algorithms on FlickerLogos-27 and our own dataset. We can see that MABO of our proposed algorithm is slightly worse (approximately 3% lower) than SS. However, we significantly reduce SS proposals by nearly 50%, which can remove some outliers and speed up trademark detection. Although Edge box achieves the best MABO on both FlickerLogos-27 and our dataset, its proposal number is tremendous compare to our proposed algorithm, which affects the performance in trademark detection.

Table 1. The MABO and number of proposals given by different algorithms on FlickerLogos-27 and our own dataset. The best results are shown in bold

Method	FlickerLogos-27 dataset		Our own dataset	
	MABO	Number	MABO	Number
BING	0.65	838.30	0.55	765.94
Edge Box	**0.74**	2363.88	**0.72**	4211.39
SS	0.69	149.96	0.67	113.30
Ours	0.67	**74.34**	0.64	**60.45**

3.4 Detection Results of Proposed Method

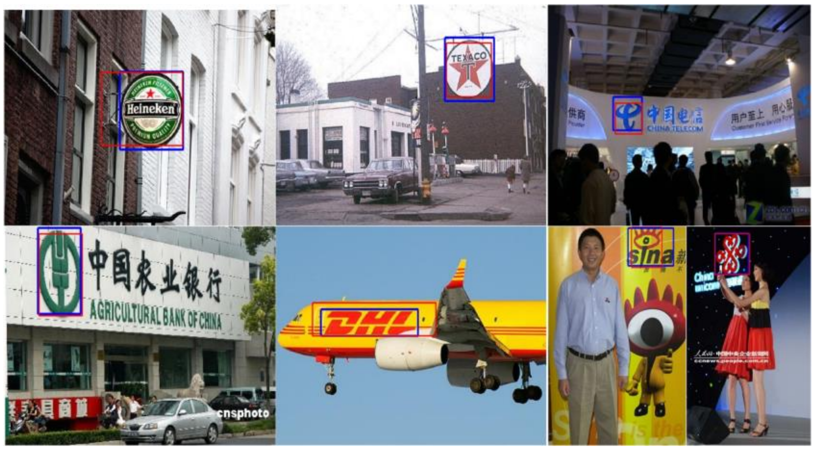

Figure 4. Examples of trademark detection on our own and FlickrLogos-27 datasets. The blue boxes are the ground truth. The red boxes are created using our method.

For a given trademark class in the dataset, we train a Logistic classifier for detection. Then, for each proposal in the image, Logistic function returns a confidence score corresponding to a specific object class. In the test dataset, every image contains one instance of the trademark. Thus, we return top-1 proposal as the result. Figure 4 indicates the reasonable detection results of the proposed method, which demonstrate that the proposed approach can provide valuable proposals for detection.

4. CONCLUSION

In this paper, we propose a novel and effective algorithm to provide high quality possible trademark locations, which integrates the MSERs and selective search. We extract adaptive MSERs from images, and then design the TCS for each MSER via Logistic regression model. Based on TCS and SS algorithm, we propose a window refinement algorithm to provide high quality proposals. Experiments on the FlickrLogos-27 and our own datasets show that the proposed algorithm can produce less but more valuable candidate trademark proposals than other algorithms, which can increase the accuracy and speed of trademark detection.

ACKNOWLEDGEMENT

This work has been supported by the National Key R&D Program of China under Grant NO.2017YFB1401000.

REFERENCES

[1] I.A. Niaz S.Y. Arafat, S.A. Husain and M. Saleem, "Logo detection and recognition in video stream," in Digital Information Management (ICDIM), 2010 Fifth International Conference on. IEEE, pp. 163-168, (2010)

[2] Yannis Kalantidis, Lluis Garcia Pueyo, Michele Trevisiol, Roelof van Zwol, and Yannis Avrithis, "Scalable triangulation-based logo recognition," in Proceedings of the 1st ACM International Conference on Multimedia Retrieval. ACM, pp. 1-7, (2011)

[3] Hang Su, Xiatian Zhu, and Shaogang Gong, “Deep learning logo detection with data expansion by synthesising context,” in IEEE Winter Conference on Applications of Computer Vision, pp. 530–539, (2017)

[4] Yuan Liao, Xiaoqing Lu, Chengcui Zhang, Yongtao Wang, and Zhi Tang, “Mutual enhancement for detection of multiple logos in sports videos,” in IEEE International Conference on Computer Vision, pp. 4856–4865, (2017)

[5] Hang Su, Shaogang Gong, and Xiatian Zhu, “Scalable deep learning logo detection,” arXiv: Computer Vision and Pattern Recognition, (2018)

[6] JRR Uijlings, KEA van de Sande, T Gevers, and AWM Smeulders, "Selective search for object recognition," International journal of computer vision, vol. 104, no. 2, pp. 154-171, (2013)

[7] Simone Bianco, Marco Buzzelli, Davide Mazzini, and Raimondo Schettini, “Deep learning for logo recognition,” Neurocomputing, vol. 245, no. C, pp. 23–30, (2017)

[8] Jiri Matas, Ondrej Chum, Martin Urban, and Tomas Pajdla, " Robust wide-baseline stereo from maximally stable extremal regions," Image and vision computing, vol. 22, no. 10, pp. 761-767, (2004)

[9] Huizhong Chen, Sam S Tsai, Georg Schroth, David M Chen, Radek Grzeszczuk, and Bernd Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, pp. 2609-2612, (2011)

[10] Lukas Neumann and Jiri Matas, "A method for text localization and recognition in real-world images," in Computer Vision-ACCV 2010, pp. 770-783. Springer, (2010)

[11] Lukas Neumann and Jiri Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, pp. 687-691,(2011)

[12] David G Lowe, "Distinctive image features from scale-invariant keypoints,"International journal of computer vision,vol.60,no.2, pp.91-110, (2004)

[13] Yuan Zhang Wei Liang, Shuwu Zhang and Qinzhen Guo, "Individualized matching based on logo density for scalable logo recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 4324-4328, (2014)

[14] C. Lawrence Zitnick and Piotr Dollr, "Edge boxes: Locating object proposals from edges,"in ECCV, (2014)

[15] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip H. S. Torr, "Bing: Binarized normed gradients for objectness estimation at 300fps," in CVPR, (2014).