Avrupa Bilim ve Teknoloji Dergisi
Sayı 27, S. 386-391, Kasım 2021

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 27, pp. 386-391, November 2021

Research Article

Brand Analysis in Social Networks Using Deep Learning Techniques

Fatma Gulsah Tan[1], Mehmet Erkan Yuksel[2*]

[1] Suleyman Demirel University, Faculty of Engineering, Department of Computer Engineering, Isparta, Turkey, (ORCID: 0000-0002-2748-0396), gulsahtan@sdu.edu.tr

[2*] Mehmet Akif Ersoy University, Faculty of Engineering and Architecture, Department of Computer Engineering, Burdur, Turkey, (ORCID: 0000-0001-8976-9964), erkanyuksel@mehmetakif.edu.tr

(First received 17 May 2021 and in final form 12 September 2021)

(**DOI:** 10.31590/ejosat.938604)

ATIF/REFERENCE: Tan, F. G., Yuksel, M. E. (2021). Brand Analysis in Social Networks Using Deep Learning Techniques. *European Journal of Science and Technology*, (27), 386-391.

Abstract

In recent years, the importance of social media data has increased with the developments in information and communication technologies, and data volume, velocity, variety, veracity, and value have been affected by these developments. Because of the popularity of social networks, the analysis of social media data has also become an important issue for large companies whose brand identity is very crucial. User comments, shares, and explanations in social networks can be used to obtain information about the brand and product. Besides, deep learning techniques, which have become popular recently and provide high accuracy, can be employed for big data analysis in social networks. The number of studies examining the brand image in social networks is quite limited. In this context, we developed a model that performs brand analysis using deep learning techniques in social networks by considering the Starbucks Coffee Company, one of the world's largest coffeehouse chains. We trained our model with Faster Region-based Convolutional Neural Network (Faster R-CNN), Single Shot Multibox Detector (SSD), Mask R-CNN, and You Only Look Once (YOLO) algorithms. We then tested the model on data from Instagram and compared the results. In the light of our results, we have shown that analyzes using deep learning techniques in social networks can significantly affect the image of companies and their brands.

Keywords: Brand analysis, Social networks, Big data, Deep learning, Convolutional neural networks.

Derin Öğrenme Teknikleri Kullanarak Sosyal Ağlarda Marka Analizi

Öz

Son yıllarda bilgi ve iletişim teknolojilerindeki gelişmelerle birlikte sosyal medya verilerinin önemi artmış, veri hacminin yanısıra veri artış hızı, çeşitliliği, doğruluğu ve değeri bu gelişmelerden etkilenmiştir. Sosyal ağların popülaritesi nedeniyle, sosyal medya verilerinin analizi marka kimliği çok önemli olan büyük şirketler için kritik bir konu haline gelmiştir. Marka ve ürün hakkında bilgi edinmek için sosyal ağlardaki kullanıcı yorumlarından, paylaşımlarından ve açıklamalarından faydalanılabilir. Buna ilaveten, son zamanlarda popüler hale gelen ve yüksek doğruluk sağlayan derin öğrenme teknikleri sosyal ağlarda büyük veri analizi için kullanılabilir. Sosyal ağlarda marka imajını inceleyen araştırma sayısı oldukça sınırlıdır. Bu kapsamda, dünyanın en büyük kahve firmalarından biri olan Starbucks örneği ele alınarak sosyal ağlarda derin öğrenme tekniklerini kullanarak marka analizi yapan bir model geliştirdik. Modelimizi Faster Region-based Convolutional Neural Network (Faster R-CNN), Single Shot Multibox Detector (SSD), Mask R-CNN ve You Only Look Once (YOLO) algoritmaları ile eğittik, Instagram sosyal medya verileri üzerinde test ettik ve sonuçları karşılaştırdık. Elde edile sonuçlar ışığında, sosyal ağlarda derin öğrenme teknikleri kullanan analizlerin, şirketlerin ve markaların imajını önemli ölçüde etkileyebileceğini gösterdik.

Anahtar Kelimeler: Marka analizi, Sosyal ağlar, Büyük veri, Derin öğrenme, Evrişimli sinir ağları.

[*] Corresponding Author: erkanyuksel@mehmetakif.edu.tr

1. Introduction

Brand analysis is a document that helps companies better understand their values. Research in a brand analysis will guide the brand to the right target and provide knowledge of the needs. With the changes brought by technology, the situation of the brands in the market is starting to deteriorate. As a result, the increase in competition creates the need for companies to try different methods to raise awareness of competitiveness (Boo et al., 2009). While creating this awareness, concepts such as brand identity and brand personality have emerged. Using social media to build our brand offers numerous opportunities to take advantage.

The rapidly increasing amount of real-world data in the social media environment has social and marketing values for large companies and government agencies, whose brand identity is vital. Powerful techniques are needed to collect data from the social media environment and perform content analysis for the data. The data used in the analysis phase should be clean, quality, and relevant to the subject. The data obtained from social media is large-volume, complex structure, and dirty. The traditional keyword-based approach to collect data about the target brand is very inadequate. The fact that images are more effective, enjoyable, and memorable than texts plays a significant role in social media branding. Images are more likely to be shared than texts. Therefore, using a versatile brand tracking method that collects relevant data based not only on the keywords developed but also on the visual content will help to overcome these problems (Dixon, 2020).

Brands have no choice but to evolve after developments in artificial intelligence. Therefore, companies should have a consumer layer by considering customers' personal information, shares, and comments. The purpose of using deep learning is to create a category-based product-level taxonomy to understand how consumers perceive brands. In this way, it helps to identify growth and development opportunities related to the relevant brand and understand how the product is perceived over time.

Deep learning can successfully analyze large amounts of data in many research fields. In this study, we developed a model that uses deep learning techniques to analyze a brand from the photos shared on social media. We chose Starbucks (one of the world's largest coffeehouse chains) as the company brand and Instagram (the most popular photo and video sharing social network service) to test our brand analysis model on social networks. Brand labeling is done by applying deep learning methods that recognize brand identity to photos taken in real-time via social media. Our study aims to search and discover the values of the relevant brand by tagging shared photos.

2. Related Work

Denton et al. (2015) presented three different hashtag embedding models that predict highly diverse and relevant hashtags for real-world Facebook images. They employed user hashtags to capture the description of image content. They made use of valuable contextual information about the user. They showed how user metadata combined with image features derived from a convolutional neural network can be used to perform hashtag prediction. They explored two ways of combining these heterogeneous features into a learning framework: simple concatenation and a 3-way multiplicative gating, where the image model is conditioned on the user metadata. They applied their models to a large dataset of de-identified Facebook posts and demonstrated that modeling the user can significantly improve the tag prediction quality over current state-of-the-art methods. Kanna et al. (2020) proposed a new intelligent methodology using deep learning to detect criminals from criminal surveillance cameras, based on the identification marks provided by the witness. This method takes gender, shirt pattern, and glasses usage as inputs to find the object as a person from the video data. The performance of this method provided an accuracy of 87% in identifying the person in the video frame. Li et al. (2020) developed a deep learning-based video detection system with a custom backbone for detecting rice diseases and pests symptoms. The system contains a frame extraction module, a still-image detector, and a video synthesis module. They used still images from their dataset to train a still-image neural network model. They employed the model to detect video frames. Their system extracts frames from video, sends the frame to the still-image detector for detection, and synthesizes the detected frame into video. It can detect multiple kinds of lesion spots in one video. The authors used Faster-RCNN as the framework in the still-image detector. They designed a custom deep convolutional neural network (DCNN) backbone of Faster-RCNN. Finally, they compared the backbones and the custom DCNN on rice diseases and pests video detection. They also proposed a set of video-based evaluation metrics based on a machine learning classifier, which reflected the quality of video detection effectively in the experiments. Nguyen et al. (2017) presented a real-time online image processing pipeline that comprises de-duplication and relevancy filtering mechanisms to collect and filter social media image content in real-time during a crisis event. They used a transfer learning approach based on state-of-the-art deep neural networks to filter out irrelevant image content. They employed perceptual hashing techniques for image de-duplication. They also performed extensive experimentation on several real-world disaster datasets to demonstrate the utility of their method. Perez et al. (2016) designed and developed deep learning-based methods to extract discriminative spatiotemporal characteristics for filtering pornographic content in videos. Their study is based on the premise that incorporating motion information in the models can alleviate the problem of mapping skin exposure to pornographic content, and advances the bar on automated pornography detection by using motion information and deep learning architectures. They presented novel ways for combining static (picture) and dynamic (motion) information using optical flow and MPEG motion vectors. They showed that both methods provide equivalent accuracies, but that MPEG motion vectors allow a more efficient implementation. Finally, they discussed results on a larger and more challenging dataset. Porzi et al. (2020) addressed and provided two important contributions for the novel task of multi-object tracking and segmentation (MOTS). They introduced an automated pipeline for extracting high-quality training data from generic street-level videos to overcome the lack of MOTS training data, without time- and cost-intensive, manual annotation efforts. The authors presented a deep-learning based MOTSNet architecture to be trained on MOTS data, exploiting a novel mask-pooling layer that guides the association process for detections based on instance segmentation masks. They provided exhaustive ablations for both their novel training data generation process and MOTSNet. Wang et al. (2016) introduced a novel learning-based logo detection method with social network information assistance.

They analyzed and used social network content to do scanning window filtering before logo detection. They proposed to utilize multiple microblog properties to perform scanning window filtering to improve detection precision and efficiency. They ranked how much a microblog relates to a brand based on multiple microblog properties. The authors presented a new dense histogram feature with spatial information integrated to classify logo and non-logo image patches in the detection stage. They conducted the experiments on Sina Weibo data. Zhao et al. (2016) proposed to predict the personalized emotion perceptions of images for each viewer. They investigated different factors that may affect personalized image emotion perceptions, including visual content, social context, temporal evolution, and location influence. They presented rolling multi-task hypergraph learning to consistently combine these factors and designed a learning algorithm for automatic optimization. For evaluation, they set up a large-scale image emotion dataset from Flickr.

3. Material and Method

In this study, we employed deep learning techniques for brand analysis. The system needs to be trained for the learning process to be carried out with the highest success. We used various CNN models to detect and classify the relevant brand in images and videos taken from the Instagram social network service. We used Faster R-CNN, Mask R-CNN, SSD, and YOLO models for our object detection system and compared the performances of these models. Figure 1 shows the application steps of our study.

Figure 1. Application stages

3.1. Faster R-CNN

It is a deep neural network method developed for reducing the burden of the region recommendations step found in previous models. It enables a unified, deep learning based object detection system to run at near real-time frame rates. It has two modules: a deep fully convolutional network that proposes regions and the Fast R-CNN detector that uses the proposed regions. The learned RPN improves the region proposal quality and the overall object detection accuracy. Thanks to the sharing of convolutional layers, it performs object detection processes with high accuracy (Huang et al., 2017; Ren et al., 2016).

3.2. Mask R-CNN

It is a simple to implement, flexible, and general framework for object instance segmentation. It efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. Mask R-CNN extends Faster R-CNN by adding a branch for predicting segmentation masks on each region of interest in parallel with the existing branch for classification and bounding box regression (He et al., 2017).

3.3. SSD

It is a fast single-shot object detector using a single DNNdeep neural network. It employs a feed-forward CNN that generates a fixed set of default bounding boxes and scores for the presence of object class examples in the boxes, followed by a non-maximum suppression step to generate final detections. SSD uses a small convolutional filter for predicting object categories and offsets in bounding box locations, using separate filters for different aspect ratio detections, and applying the filters to multiple feature maps from the later stages of a network to carry out detection at multiple scales (Liu et al., 2016; Diwan et al., 2021).

3.4. YOLO

It is a real-time object detection system. It frames object detection as a regression problem, straight from image pixels to spatially separated bounding boxes and associated class probabilities. It employs a single CNN that divides the image into regions and predicts bounding boxes and class probabilities directly from full images in one evaluation. It is faster than other algorithms. However, it cannot detect a group of small objects or irregularly shaped objects within an image (Redmon et al., 2016).

3.5. Training, Validation, and Test

The four hundred images of the Starbucks brand were collected via Google Image to use in the training model. 20% of these images were used for test data and 80% for training data. The brand is individually labeled in each image. LabelImg (a tool for object labeling in TensorFlow library models) is employed for the labeling process. After the labeling process, XML files to be used in the training model were created for each image. Figure 2 shows the sample images and XML files. For the trained model to have a high performance, it is important that the images are taken from different angles, are varied, and the number of them is large.

Figure 2. Sample images and corresponding XML files

To develop the Starbucks brand detection model, existing object detection models were retrained using the data sets created. We determined which model and parameters will be used together with the labeling information of the brand object. We determined the labeling information (classification number) as starbuckslogo. We used Python programming language, TensorFlow library, faster_rcnn_inception_v2_coco, mask_rcnn_inception_v2_coco, SSD-Single Shot Multibox MobileNet_v2, and YOLOv3 models. Parameters are the same as the labeling information. The parameter number is 1, and the parameter name is starbuckslogo. We employed a Lenovo Legion Y520 computer with Intel Core i7 CPU and Nvidia GTX 1050 Ti GPU. The training time took approximately 6 hours for Faster R-CNN, Mask R-CNN, and SSD and about 3 hours for YOLO. The YOLO model was implemented in 3000 steps and the other models in 10,000 steps. The number of steps can be changed according to the selected model, the data, and the number of objects. We tested four models trained to detect the Starbucks brand on twenty different images and three videos obtained from Instagram. Figure 3, Figure 4, Figure 5, and Figure 6 show the resulting images of the model trained by Faster R-CNN, Mask R-CNN, SSD, and YOLO, respectively.

Figure 3. Results of the model trained by Faster R-CNN

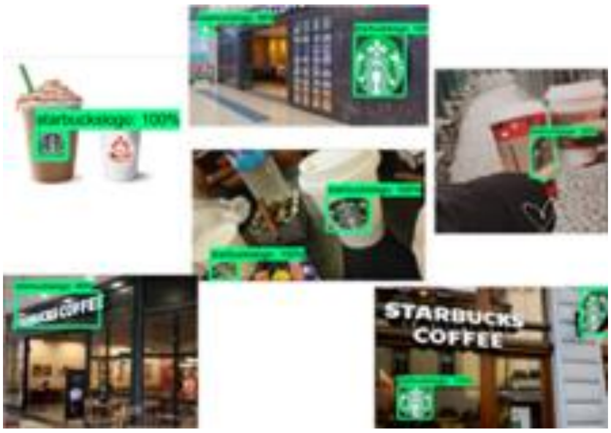

Figure 4. Results of the model trained by Mask R-CNN

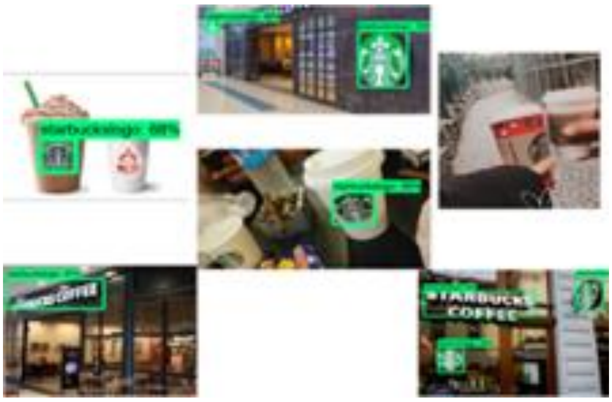

Figure 5. Results of the model trained by SSD

Figure 6. Results for the model trained by YOLO

We observed that the fastest model in the training and testing phases is YOLO. The training process was completed in 3000 iterations according to the missing graphic values. In other models, the speed order is as follows: SSD, Faster R-CNN, and Mask R-CNN. In these models, the training process was completed in 10000 iterations. YOLO performed the object detection by processing 26 seconds of video in about 10 seconds, SSD in 1 minute, Faster R-CNN, and Mask R-CNN in 8 minutes.

4. Results and Discussion

The variety of information used for object detection increases the model success however causes an increase in processing time. Object detection is slow since models developed before Faster R-CNN apply more than one operation to the data. In this study, training models were created with Faster R-CNN, Mask R-CNN, SSD, and YOLO models, and their results were compared since the speed and consistency ratio are important in object detection. We employed widely accepted validation metrics to evaluate the performance of deep learning techniques. SSD can perform object recognition faster than the Faster R-CNN and Mask R-CNN models since it makes only one forward movement. Considering the consistency rate, the success rate of Faster R-CNN and Mask R-CNN is higher than the SSD. Mean Average Precision (mAP) was used to calculate the accuracy of the model. The mAP is a key performance metric in many multi-class classification tasks and is used to measure the accuracy of object detectors. It calculates the average precision value for the recall value from 0 to 1.

Precision is a measure of how much is correctly predicted from all classes. It should be as high as possible. It is the ratio of True Positive (TP) examples divided by the sum of True Positive (TP) and the False Positive (FP) examples.

$$Precision = TP/(TP + FP) \tag{1}$$

Recall measures the percentage of all relevant data returned by the classifier. It means that the model returns most of the relevant data.

$$Recall = TP/(TP + FN) \tag{2}$$

Average Precision (AP) is calculated by averaging the precision of the individual recall values for each class. The average of the AP values gives the MAP value. Table 1 shows the mAP values.

Table 1. mAP values

Model	mAP
Faster R-CNN	0.96
Mask R-CNN	0.96
SSD	0.76
YOLO	0.98

The average sensitivity for object detection by Faster R-CNN was 0.96, 0.96 for Mask R-CNN, 0.76 for SSD, and 0.98 for YOLO. When we compared the total loss of Mask R-CNN, SSD, and YOLO, we observed that Faster R-CNN starts from 1.1 and approaches below 0.1. We obtained similar results for Mask R-CNN. The loss of SSD in 10000 steps is close to 1, starting from 11.7. For YOLO, we observed the loss started from 1 and approached below 0.1 at the 1000th step. The results show that YOLO has higher performance than Faster R-CNN, Mask R-CNN, SSD, and Faster R-CNN and Mask R-CNN are stronger than SSD. Figure 7 depicts the loss curve plot of Faster R-CNN, Mask R-CNN, SSD, and YOLO models.

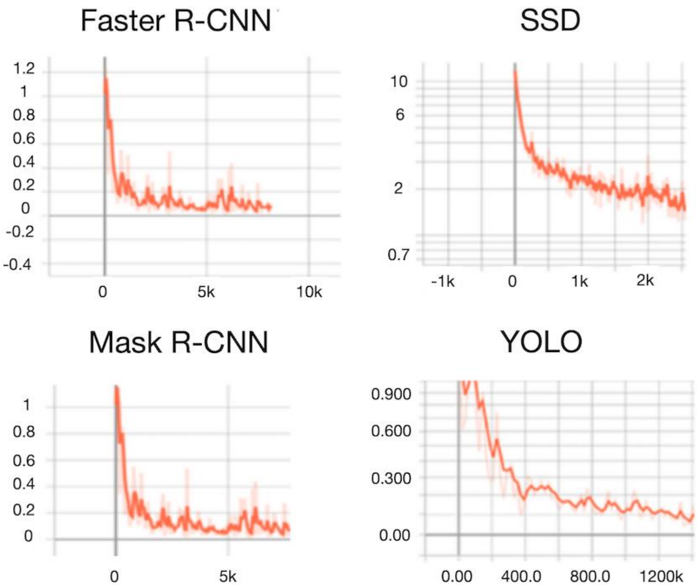

Figure 7. Loss curves of deep learning models

We used the visual logos of Starbucks in the image labeling. We determined the number of classes as one. We did not use the text logo because it can be in different languages. Since the labeling process was not performed, the results obtained during the test phase were ignored. The text logo of Starbucks must be evaluated in a different class, and labeling must be performed on the data belonging to it. This procedure helps to make better brand analysis by increasing the performance of the training model used.

Increasing images on social networks pose both a threat and an opportunity for companies' brands. With the applications developed by deep learning techniques, analyzes can be made on these social platforms, brands can be better understood, customers can be better known, and marketing strategies can be developed efficiently. Using scalable deep learning methods, it becomes possible to analyze billions of images and videos per day. In this study, we tested the training model with the data on Instagram, and we showed that this model can be employed for brand analysis easily in social networks. We used only videos and pictures in the brand analysis. Models can be developed using comments and tags under images, and better analyzes can be performed. Using a large data set can enable the realization of social predictions more effectively. Thus, models developed can contribute to company decision-making processes for effective management plans.

5. Conclusions and Recommendations

In this study, we employed CNN-based techniques for brand analysis and real-time object detection in social networks. Faster R-CNN, Mask R-CNN, SSD, YOLO methods, and a dataset of Starbucks brand were used in the training model. Training time took approximately three hours for the YOLO model and about six hours for the other models. Video and image datasets collected from Instagram were used to analyze the model performance. YOLO can be preferred if both consistency rate and speed are essential in the applications. SSD works fast without the need for high-level hardware, but it cannot make accurate detections compared to other models. Faster R-CNN and Mask R-CNN create more accurate models than SSD. In our tests, we observed that the YOLO has a higher performance in real-time object detection. Consequently, we showed that the model we developed can be used on every platform for every brand and can easily perform brand analysis by using it on the stacked dataset.

References

Boo, S., Busser, J., & Baloglu, S. (2009). A model of customer-based brand equity and its application to multiple destinations. *Tourism Management*, 30, 219-223. https://doi.org/10.1016/j.tourman.2008.06.003

Denton, E., Weston, J., Paluri, M., Bourdev, L., & Fergus, R. (2015). *User conditional hashtag prediction for images*. The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, Sydney, NSW, Australia. https://doi.org/10.1145/2783258.2788576

Diwan, A., Gupta, V., & Chadha, C. (2021). Accident detection using mask R-CNN. *International Journal for Modern Trends in Science and Technology*, 7(01), 69-72.

Dixon, A. (2020). Efficient clustering for users' brand sentiment analysis on online social media. University Presentation Showcase Event.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. IEEE International conference on computer vision, Venice, Italy.

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., & Murphy, K. (2017). *Speed/accuracy trade-offs for modern convolutional object detectors*. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA. https://doi.org/10.1109/CVPR.2017.351

Kanna, J. S. V., Raj, S. M. E., Meena, M., Meghana, S., & Mansoor Roomi, S. (2020). *Deep learning based video analytics for person tracking*. International Conference on Emerging Trends in Information Technology and Engineering, Vellore, India. https://doi.org/10.1109/ic-ETITE47903.2020.173

Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., & Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors*, 20(3), 578. https://doi.org/10.3390/s20030578

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *Lecture notes in computer science: Vol. 9905. Computer Vision-ECCV 2016* (pp. 21-37). Springer. https://doi.org/10.1007/978-3-319-46448-0_2

Nguyen, D. T., Alam, F., Ofli, F., & Imran, M. (2017). Automatic image filtering on social networks using deep learning and perceptual hashing during crises. arXiv: 1704.02602.

Perez, M., Avila, S., Moreira, D., Moraes, D., Testoni, V., Valle, E., & Rocha, A. (2016). Video pornography detection through deep learning techniques and motion information. *Neurocomputing*, 230, 279-293. https://doi.org/10.1016/j.neucom.2016.12.017

Porzi, L., Hofinger, M., Ruiz, I., Serrat, J., Bulo, S., & Kontschieder, P. (2020). *Learning multi-object tracking and segmentation from automatic annotations*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA. https://doi.org/10.1109/CVPR42600.2020.00688

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. The 29th IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA.

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

Wang, F., Qi, S., Gao, G., Zhao, S., & Wang, X. (2016). Logo information recognition in large-scale social media data. *Multimedia Systems*, 22, 63-73. https://doi.org/10.1007/s00530-014-0393-x

Zhao, S., Yao, H., Gao, Y., Ji, R., Xie, W., Jiang, X., & Chua, T. (2016). *Predicting personalized emotion perceptions of social images*. The 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands. https://doi.org/10.1145/2964284.2964289