Advertisement and Document Recommendation based on Content in the Image

Meenakshi Chandak
ME Computer
PICT, Dhankawadi, Pune, Maharashtra, India- 411013

A. S. Ghotkar, PhD
Professor, Computer
PICT, Dhankawadi, Pune, Maharashtra, India- 411013

ABSTRACT
In recent years, more and more images have been uploaded and published on the web. Along with text web pages, images have become an important media for various social media platforms to place relevant advertisements. However, conventional image advertising primarily uses text content rather than image content to match relevant advertisements. There is no existing system to automatically monetize the opportunities brought by individual image. As a result, the advertisements are only generally relevant to the entire web page rather than specific to images it contained. To overcome this, advertisements in the proposed system are recommended based on images. The objects are detected from the image using TensorFlow API Model and based on those objects (keywords) advertisements are recommended. An additional application is provided, were based on the detected objects (keywords) relevant documents are recommended using Term Frequency-Inverse Document Frequency algorithm. From the experimental results, it is seen that system could recognize over 90 percent of objects and could recommend relevant advertisement with mean average precision of 0.66.

General Terms
Computer Vision, Image Processing

Keywords
Advertisement Recommendation, Document Retrieval, Object Detection, Topic Modeling.

1. INTRODUCTION
Advertising has embarked on a dramatic evolution, which will be rapid, fundamental and permanent. Although this evolution is still underway in advertising in terms of objectives, strategy, and solutions, we can summarize the trends of Internet advertising into two generations in terms of methodologies: conventional advertising and contextual advertising.

Nowadays, web pages no longer contain just textual information. Instead, more and more images have been uploaded and published on the web. For instances, social web sites like Facebook and Flickr have billions of photo album pages with little text. Compared with the traditional textual web pages, images become the main contents of these web pages. Thus, traditional contextual advertising approaches cannot be directly applied to web pages dominated by images because of the lack of textual information. Therefore, understanding the contents or topics of images and then recommending relevant advertisements based on these images becomes a challenging problem, interesting to both academia and industry.

Considering the problem of recommending advertisements for web images, traditional methods largely rely on the textual contexts of images, such as surrounding text and social tags, to extract keywords and then obtain relevant advertisements through textual information retrieval. However, there are a large amount of web images with little or no text contexts. Furthermore, text can be noisy and ambiguous, which could reduce the accuracy for the recommended advertisements. Ideally, in order to perform visual contextual advertising, the algorithm must first understand the images and then make appropriate recommendations based on the understanding.

When a user enters a search query, or browses a web page, or more generally, interacts with some text, the advertisement platform will select and show relevant advertisements based on the text content in the query or the page. Though other contextual information, such as location, time, and user profile can be taken into consideration, textual understanding is still the primary technology here. This approach however neglects a significant amount of potential user interest available within images. With the recent emergence of online multimedia services, people can easily share photos with their friends in real time. How to retrieve user interests from the content of images, and provide an advertisement platform based on image bidding, has therefore become a promising research direction. Many existing ad-networks such as Google AdSense, Yahoo, and BritePic have provided contextual advertising services around Images.

The proposed method is an image-centric advertisement platform. In this platform, images are the main input. The objects are detected from the image using TensorFlow API model and based on detected objects advertisements are recommended. An additional application, based on the detected objects (keywords) relevant documents are recommended using Tf-Idf algorithm.

2. RELATED WORK
This chapter includes approaches to advertisement recommendation system driven by images. We studied these approaches which helped us to resolve advertisement recommendation system problem. These approaches are discussed in this chapter. T. Mei, S. Hua, et al. proposed an innovative contextual advertising system driven by images, which automatically associates relevant advertisements with an image rather than the entire text in a web page. The proposed system, called ImageSense, represents the first attempt towards Contextual In-Image Advertising. The relevant advertisements are selected based on not only textual relevance but also visual similarity so that the advertisements yield contextual relevance to both the text in the web page and the image content. [1]

W. Jiang, D. Liu, et al. proposed a new advertisement platform which allows search engine advertisers to bid on images instead of just plain text. The main components of this platform include an advertisement editorial tool, ROI detection, image content understanding, and image matching modules. This platform is suitable for application scenarios where images are the main input, for example, in Multimedia Messaging Service (MMS) or content-based image retrieval. [2]

Y. Chen, O. Jin, et al. addressed the novel problem of visual contextual advertising, which is to directly advertise when users are viewing images which do not have any surrounding text. A key challenging issue of visual contextual advertising is that images and advertisements are usually represented in image space and word space respectively, which are quite different with each other inherently. As a result, existing methods for webpage advertising are inapplicable since they represent both web pages and advertisement in the same wordspace. In order to solve the problem, authors proposed to exploit the social web to link these two feature spaces together. In particular, authors presented a unified generative model to integrate advertisements, words and images. Specifically, solution combines two parts in a principled approach: First, transform images from an image feature space to a word space utilizing the knowledge from images with annotations from social web. Then, a language model based approach is applied to estimate the relevance between transformed images and advertisements. Moreover, in this model, the probability of recommending an advertisement can be inferred efficiently given an image, which enables potential applications to online advertising. [3]

D. V. Phuong, T. M. Phuong, et al. Proposed a method which uses a keyword topic model that associates each keyword provided by the advertiser with a multinomial distribution over topics unlike existing methods that directly model the content of an advertisements as a distribution over topics. Then, an advertisement with multiple keywords is represented as a mixture of topic distributions associated with those keywords. [4]

Y. Kalantidis, A. Farahat, et al. proposed a new system for selecting and displaying visual advertisements in image search result sets. The method compares the visual similarity of candidate advertisements to the image search results and selects the most visually similar advertisement to be displayed. [5]

A. Broder, M. Fontoura, et al. proposed a novel way of matching advertisements to web pages that rely on a topical (semantic) match as a major component of the relevance score. The semantic match relies on the classification of pages and advertisements into a 6000 nodes commercial advertising taxonomy to determine their topical distance. As the classification relies on the full content of the page, it is more robust than individual page phrases. The semantic match is complemented with a syntactic match and the final score is a convex combination of the two sub-scores with the relative weight of each determined by a parameter. [6]

C. Xiang, T. V. Nguyen, et al. proposed a novel advertising technique called SalAd, which utilizes textual information, visual content and the web page saliency, to automatically associate the most suitable companion advertisements with online videos. SalAd consists of three basic steps. Given an online video and a set of advertisements, first roughly identify a set of relevant advertisements based on the textual information matching. Then carefully select a sub-set of candidates based on visual content matching. In this regard, selected advertisements that are contextually relevant to online video content in terms of both textual information and visual content. Finally select the most salient advertisement among the relevant advertisements as the most appropriate one. [7]

3. METHODOLOGY

The proposed system shown in Figure 1 takes image as the input (which contains various objects in it) and detects the object from the image using TensorFlow API Model. Detected objects are used as keywords for advertisement recommendation as well as for Document Retrieval. For advertisement recommendation BeautifulSoup libraries are used. More the objects detected from the image more relevant advertisement and document can be recommended.

Why SSD (Single Shot Detection) Mobilenet TensorFlow API model? For object detection, among all the TensorFlow API model, the one used in this project is SSD Mobilenet Model as it is lightweight and fastest. It is clear from Figure 2 that higher the mAP (mean average precision) score, the more accurate the model is but that comes at the cost of execution speed.

The network for TensorFlow API used in this project is based on Single Shot Detection (SSD). The architecture used in this project is with input 300x300x3. The SSD normally starts with a VGG model, which is converted to a fully convolutional network. The output at the VGG network is a 38x38 feature map (conv4x3). The added layers produce 19x19, 10x10, 5x5, 3x3, 1x1 feature maps. All these feature maps are used for predicting bounding boxes at various scales.

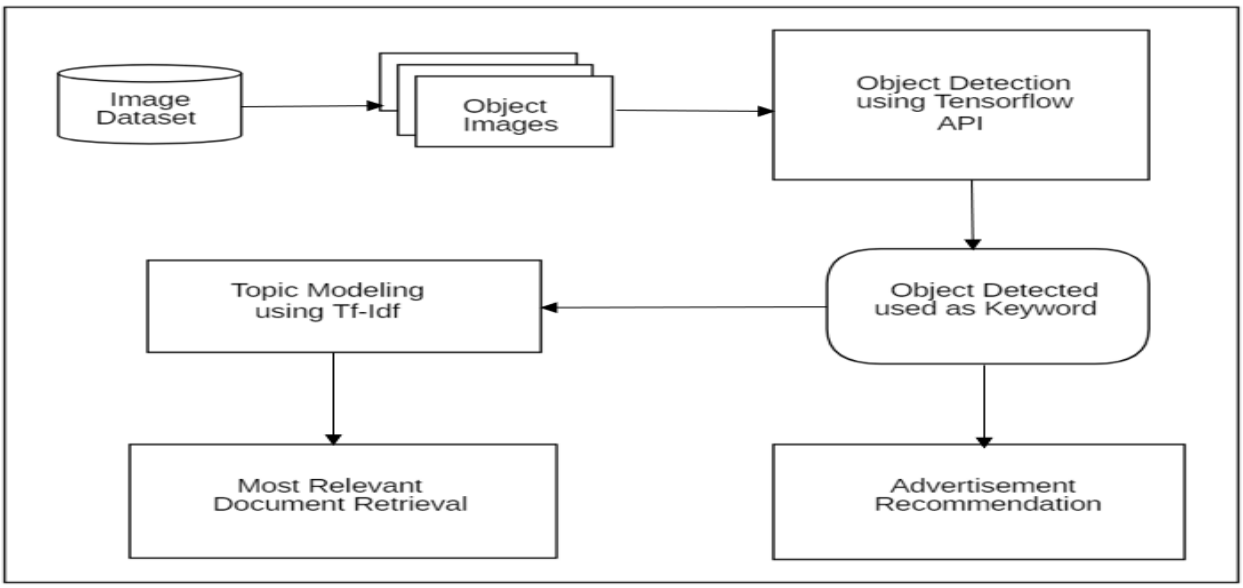

Figure 1: Proposed System Diagram for Advertisement Recommendation and Document Retrieval Based on Image Content.

Model name	Speed	COCO mAP	Outputs
ssd_mobilenet_v1_coco	fast	21	Boxes
ssd_inception_v2_coco	fast	24	Boxes
rfcn_resnet101_coco	medium	30	Boxes
faster_rcnn_resnet101_coco	medium	32	Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	slow	37	Boxes

Figure 2: Comparison of TensorFlow API Model [14]

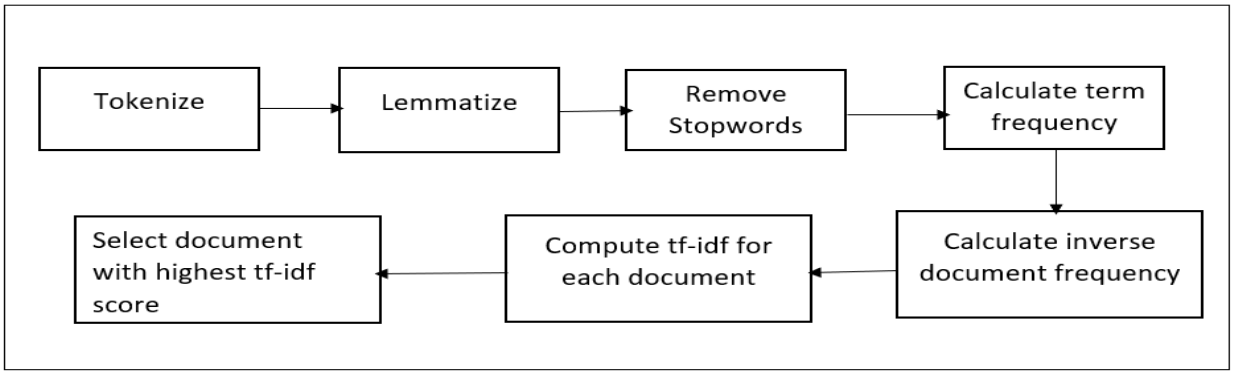

Figure 3: VVG-16 Architecture [14]

VVG-16 Architecture consist of:

1. Convolutions layers
2. Max pooling layers
3. Fully connected layers at end
4. Total 16 layers

Figure 3 explains steps for how to retrieve relevant document from dataset. The algorithm used for document retrieval is Tf-Idf. The steps are as follows:

1. First step includes tokenizing each document.

2. Lemmatize each document (to find keywords that appear in different forms like cars, car's).

3. Removal of stopwords like is, an, the, a and soon.

4. For every term in the current document, and every document in the set, compute the term frequency (how many times the term occurs in the document).

5. For every term in the current document, and every document in the set, compute the inverse document frequency.

6. For every term in the current document, and every document in the set, compute the Tf-Idf score.

7. Select the document with highest Tf-Idf Score.

4. RESULTS AND ANALYSIS

Figure 4 shows graphical representation of object detection accuracy. Here accuracy for detecting each object is shown separately. Accuracy for detecting objects can be calculated as follows:

$$Accuracy\_of\_objects\_detected = \frac{\sum_{i=1}^{n} Correctly\_detected\_objects}{Total\_number\_of\_images} \times 100$$

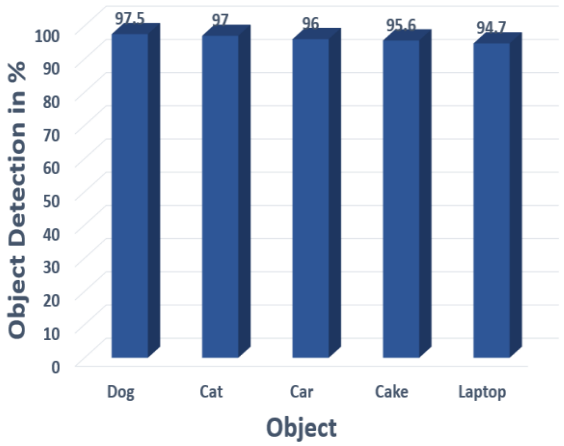

Figure 4: Object Detection Accuracy

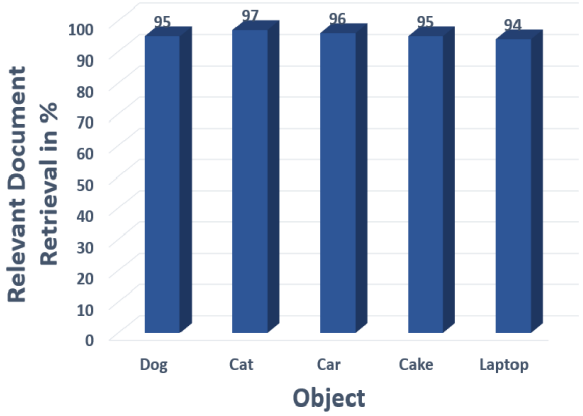

Figure 5: Document Retrieval Accuracy

Figure 5 shows graphical representation of relevant document retrieval accuracy. From the figure clearly that relevant document retrieval for the system is above 90%. Accuracy of relevant document retrieved can be calculated as follows.

$$Accuracy\_of\_document\_retrieval = \frac{\sum_{i=1}^{n} Relevant\_document\_retrieval}{n} \times 100$$

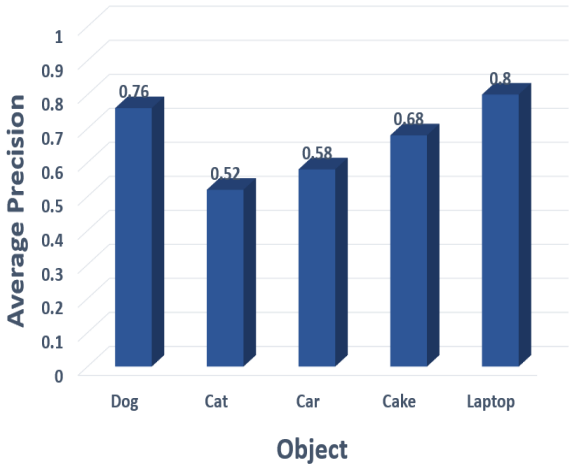

Figure 6: Average Precision

Figure 6 shows graphical representation of average precision for advertisements recommendation by different persons Precision can be calculated as follows:

$$Precision = \frac{Relevant\_Advertisements}{Retrieved\_Advertisements}$$

Table 1: Average Precision and mAP (Mean Average Precision)

Object/ Person	Dog	Cat	Car	Cake	Lap-top	Avg
Person 1	0.8	0.3	0.8	0.8	0.7	0.62
Person 2	0.8	0.6	0.6	0.6	0.8	0.68
Person 3	0.7	0.6	0.6	0.6	0.8	0.66
Person 4	0.7	0.6	0.6	0.8	0.8	0.74
Person 5	0.8	0.3	0.6	0.6	0.9	0.64
Average Precision/ mAP	0.76	0.52	0.58	0.68	0.8	0.66

Table 1 shows precision for each object on different persons opinion. Here precision is calculated by taking the opinion of each person with respect to retrieved advertisements whether they are relevant or not. Average precision and mean average precision are calculated in the table.

5. CONCLUSION

Online advertising is one of the challenging problems in computer vision and machine learning. In this project, a new approach is proposed for online advertising, called image contextual advertising, which is to recommend advertisement for an image without the help of any surrounding text. Advertisers bid on content in the images, instead of text. The object detected from the image serve as keywords for advertisement recommendation. The efficiency and usability of solution is demonstrated through experimental results. From the results, it can be seen that our system could recognize over 90% of objects and could recommend relevant advertisement with mean average precision of 0.66. An additional document retrieval application is developed based on the images.

6. REFERENCES

[1] T. Mei, X. Hua, "Contextual In-Image Advertising," Proceedings of the 16th ACM International Conference on Multimedia, pp. 439-448, 2008.

[2] W. Jiang, Dechao Liu, "An Online Advertisement Platform based on Image Content Bidding," IEEE International Conference on Multimedia and Expo, pp. 1234 - 1237, 2009.

[3] Y. Chen, O. Jin, "Visual Contextual Advertising: Bringing Textual Advertisements to Images," Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, pp. 1314-1320, 2010.

[4] D. V. Phuong, T. M. Phuong, "A Keyword-Topic Model for Contextual Advertising," Proceedings of the ThirdSymposium ACM on Information and Communication Technology, pp. 63-70, 2012.

[5] Y. Kalantidis, A. Farahat, "Visual Congruent Ads for Image Search," Proceedings of 23rd International Conference on Pattern Recognition in IEEE, pp. 1496-1505, 2016.

[6] A. Broder, M. Fontoura, "A Semantic Approach to Contextual Advertising," Proceedings of the 30th annual international ACM SIGIR Conference on Research and development in Information Retrieval, pp. 559-566, 2007.

[7] C. Xiang, T. V. Nguyen, "SalAd: A Multimodal Approach for Contextual Video Advertising," Proceeding in IEEE International Symposium on Multimedia, pp. 211-216, 2015.

[8] S. Wang, Z. Chen, "Identifying Search Keywords for Finding Relevant Social Media Posts," AAAI Conference on Artificial Intelligence, pp. 3052-3058, 2016.

[9] W. Zhang, D. Wang, "Advertising Keywords Recommendation for Short-Text Web Pages Using Wikipedia," ACM Transactions on Intelligent Systems and Technology archive, vol. 3, no. 2, pp. 3101-3136, 2012.

[10] Y. Y. Chen, T. Chen, "Predicting Viewer Affective Comments Based on Image Content in Social Media," Proceedings in ACM International Conference on Multimedia Retrieval, pp. 233-241, 2014.

[11] T. Chen, F. X. Yu, "Object-Based Visual Sentiment Concept Analysis and Application," Proceedings of the 22nd ACM International Conference on Multimedia, pp. 367-376, 2014.

[12] T. Mei, X. S. Hua, "Contextual Internet Multimedia Advertising," Proceedings of the IEEE, vol. 98, no. 8, pp. 1416 - 1433, Apr 2010.

[13] J. Sumalatha, H. Girish, "Topic Modeling using TF-IDF and Linked Data," International Journal of Engineering Research in Computer Science and Engineering, vol. 5, no. 4, pp. 2320-2394, 2018.

[14] Y. Feng and M. Lapata, "Topic Models for Image Annotation and Text Illustration," Proceeding Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 831-839, 2010.

[15] M. Abadi, M. Isard, "A computational model for TensorFlow: an introduction," Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages, pp. 1-7, 2017.

[16] J. Huang, C. Sun, "Speed/accuracy trade-offs for modern convolutional object detectors," Proceedings of Cornell University Library in Computer Vision and Pattern Recognition, pp. 7310-7319, 2016.