"A STUDY OF UNIFIED DETECTION OF TEXT AND RECOGNITION OF OBJECTS IN NATURAL SCENE IMAGES."

By
Kaushik Das

SYNOPSIS

INTRODUCTION

There is a long history of development of automated readers. The first system was patented in 1951 by Shephard, D. H [1]. This device was able to convert a few text type documents into punch cards available in a magazine subscription department. Since then, many devices and software techniques are improving thereby creating challenges to the researchers. There is a difference in reading text in documents and scene images. The primary difference is locating the text to be recognized. If we consider a one or two column simple document, then it is easy to identify the text in the document and no detection will be required. On the other hand, if we consider a complex document like newspaper or magazine, there is a challenge of distinguishing between text and image regions. Natural scene classification is a challenging problem for computer vision, since most scenes are collections of entities (e.g. objects) organized in a highly variable layout. This high variability in appearance has made flexible visual representations quite popular for this problem [2].

Object recognition is a very challenging job as it is almost a regular practice by a person starting from whether a surface is easy to walk on, or what kind of grip to use to pick up an object or to recognize objects those make up our surroundings. Automatic object recognition can be widely useful in a variety of applications, including robotics, product search, and image editing for interior design. But automatic object recognition of real world images is very challenging. The materials may come of various categories such as fabric or wood and visually very diverse in appearances. The appearance may also vary due to lightning or shape and thereby causing Object recognition as very demanding approach in present days.

LITERATURE SURVEY

In the paper title "*Scene Classification with Semantic Fisher Vectors*" authored by Mandar Dixit, Si Chen, Dashan Gao, Nikhil Rasiwasia and Nuno Vasconcelos [3] it has been mentioned that a scene image has been represented as a bag of semantics (BoS) with the help of a Convolutional Neural Network (CNN) trained to recognize objects. Recent efforts of improving scene classification have relied on a pre-trained imageNET CNN mainly because of the superior quality of its feature responses. When combined with features from a scene classification CNN, their semantic Fisher Vectors (FV) produces state-of-the-art results.

Zhou et al. [4] proposed a more direct approach that does not rely on the ImageNET CNN [17] at all. They simply learn a new CNN on a large scale database of scene images known

as the “Places” dataset [5]. Although the basic architecture of their Places CNN is the same as that of the ImageNET CNN, the type of features learned is very different.

Junjie Yan et al. in their *IEEE Xplore* paper *“Object Detection by Labeling Superpixels”,* [5] proposed to handle object detection by labeling superpixels. Compared with the traditional proposal generation and classification based methods [6], the superpixel based method has a much larger output space and provides more flexibility. It can infer overlapped objects by encoding global image information. Current leading methods, such as RCNN with very deep CNN, can be incorporated into the superpixel labeling by providing a strong data cost term.

Qingfeng Liu et al. in their IEEE Xplore paper “*A Novel Locally Linear KNN Model for Visual Recognition*” [7] presents a novel locally linear KNN model for robust visual recognition. They proposed a locally linear nearest mean classifier (LLNMC) KNN based classifier that reveals its connection to the Bayes decision rule for minimum error in the context of kernel density estimation.

Fei Yan et al. in their paper “*Deep Correlation for Matching Images and Text*” [8] addressed the problem of matching images and captions in a joint latent space learnt with deep canonical correlation analysis (DCCA).

Haoxiang Li et al. in their paper “A Convolutional Neural Network Cascade for Face Detection” [9] presented a CNN cascade for fast face detection by evaluating the input image at low resolution too.

Sean Bell et al. in their Xplore paper titled “Material Recognition in the Wild with the Materials in Context Database”[10], introduced a new, large-scale, open dataset of materials in the wild, the Materials in Context Database (MINC), and combined this dataset with deep learning to achieve material recognition and segmentation of images in the wild.

MOTIVATION

The concept of automated information retrieval (IR) was developed in the 1950s, which became an active research area in the 1960s, but only achieved mass usage in the 1990s with Web search engines. From the literature survey, it has been observed that there is enormous scope of text recognition in image. There is scope of development in the recognition mechanism so that the exact recognition of text present in an image can be done. In the image in the figure no. 1, the original image containing “NICE SKY” cab be viewed but the program execution retrieved as “N1CE SKY”. In this example the difference between 1 and I could not be distinguished. Similarly a study on other such remedies and its correction has been found necessary.

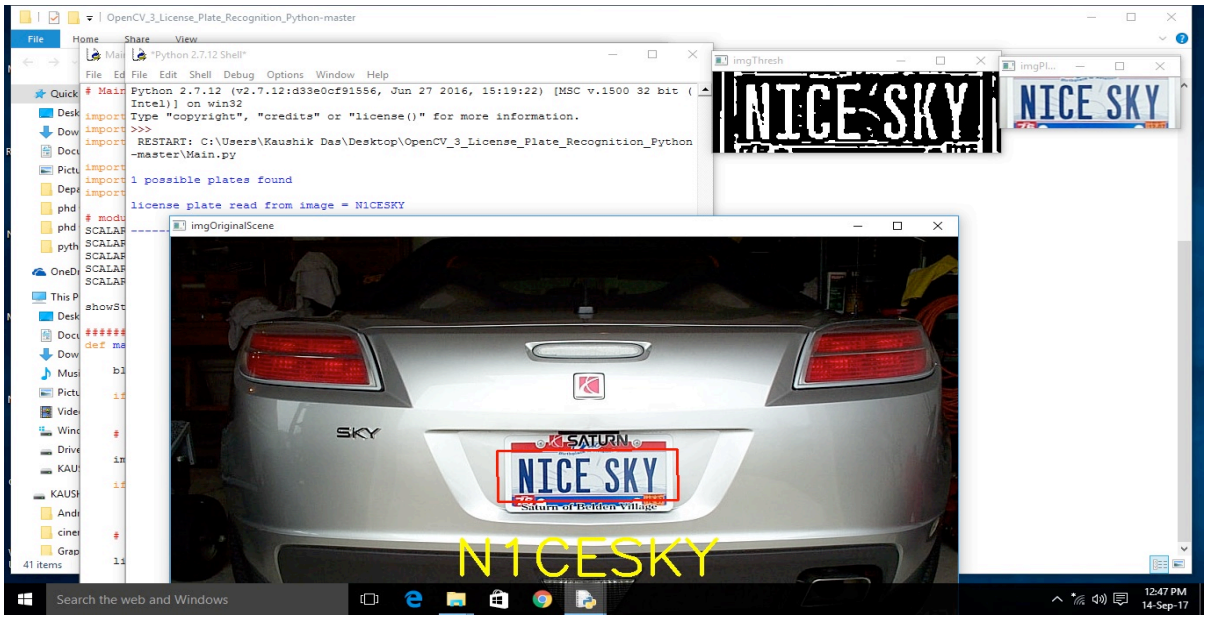

Figure no: 1

The advantages and disadvantages for using KNN are mentioned below:

Advantages:

1. It is simple to use and it has effectiveness, intuitiveness and competitive classification performance in many domains.
2. It is Robust to noisy training data
3. It is effective if the training data is large.
4. Simple and powerful. No need for tuning complex parameters to build a model.
5. It is a powerful non-parametric classification system which bypasses the problem of probability densities completely

Disadvantages:

1. Distance based learning is not clear which type of distance to use and which attribute to use to produce the best results.

2. Computation cost is quite high because we need to compute distance of each query instance to all training samples.

The ideal superpixel partition for detection is that the superpixel number is small enough for the efficiency in inference and each superpixel does not span in multiple objects. The removal of noise using superpixel is a challenging factor.

The development of applications which lead to reading contents in an image is of great need and there are a few applications which are still lacking behind retrieving exact text present in an image.

The recognition of objects is also a challenging factor in the world of digitalization. The world is developing in terms of less man power utilization by input from scanners and digitally recognition of objects. The ongoing research on recognition is on high demand in most of the universities and R&D sectors all over the world.

AIM AND OBJECTIVES

In the thesis entitled "A study of unified detection of text and recognition of objects in natural scene images" an attempt has been made to detect the text in an image and extract the data in text character using KNN algorithm with a detailed study about refinement of text detection present in an image. An attempt will be made for modification of the KNN algorithm after a detailed study of the different aspects of KNN algorithm [12, 13, 14].

It has been observed that the original images may contain a huge amount of noise. Our work will concentrate on the images with a reducible amount of noise only and focus on the removal of noise using the ideal superpixel partition for detection.

Further, study of object recognition will be done throughout our work. In the work, we will make one new contribution:

- We will introduce a new material dataset, Materials in Context Database (MINC), and 3-stage crowdsourcing pipeline for efficiently collecting millions of click labels.

The prime objectives of our study are to obtain high quality segmentation results in the detection of text in an image and recognition of objects.

METHODOLOGY

The chapterization of the proposed thesis will be made as per the objectives of the study. The first chapter of the proposed thesis will begin with the basic discussion of the pioneering KNN algorithm, scope of improvement of KNN algorithm and techniques to remove certain noise in an image containing texts. The filtered image after noise removal will be considered for text detection on image and certain basic results by using OpenCV and NumPy packages in python programming language. Moreover an analysis on the problem including its theoretical and practical aspects will also be given in this chapter. An intensive literature review on the KNN algorithm to be used for Text Detection and Recognition of objects present in an image will also be included in this chapter.

In the second chapter we will include a study on modified KNN algorithm and how it can be implemented using OpenCV and NumPy package in python. We will compare about the few modified KNN algorithms presented by different authors. The main idea for focusing on this algorithm will be presented and a comparison with other similar algorithms will also be made using same training datasets.

In the third chapter, we will concentrate on the images with a reducible amount of noise only and focus on the removal of noise using the ideal superpixel partition for detection. The images considered as standard for the text detection will be taken into account for removing noise before detection of the text. The processed image will also be discussed with proper analysis.

The fourth chapter will give a detailed study about the text detection algorithms using the improved KNN algorithm that will be discussed in second chapter. An analysis of the results will

also be made with an attempt to discover better results in comparison to some other techniques found in the literature survey.

The fifth chapter will make a study of object recognition with one new contribution by introducing a new material dataset, Materials in Context Database (MINC) [10], and 3-stage crowdsourcing pipeline for efficiently collecting millions of click labels.

In the last chapter of our thesis we will give a summary of the findings of our whole work and will also discuss the future scope of research in this area.

REFERENCES

[1] Ahmed, E., Jones, M. and Marks, T. K., 2015," *An Improved Deep Learning Architecture for Person Re-Identification*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3908-3916.

[2] Akata, Z., Reed, S., Walter, D., Lee, L. and Schiele, B., 2015," *Evaluation of Output Embeddings for Fine-Grained Image Classification*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2927-2936.

[3] Dixit, M., Chen, Si., Gao, D.,Rasiwasia, N. and Vasconcelos, N., 2015," *Scene Classification with Semantic Fisher Vectors*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2974-2983.

[4] He, K. and Sun, J., 2015," *Convolutional Neural Networks at Constrained Time Cost*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 5353-5360.

[5] Yan, J., Yu, Y., Zhu, X., Lei, Z. and Li, S. Z., "*Object Detection by Labeling Superpixels*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR).

[6] Byeon, W., Bruel, T. M., Raue, F. and Liwicki, M., 2015," *Scene Labeling with LSTM Recurrent Neural Networks*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3547-3555.

[7] Liu, Q. and Liu, C., "*A Novel Locally Linear KNN Model for Visual Recognition*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Yan, F. and Mikolajczyk, K., "*Deep Correlation for Matching Images and Text",* was published in the Conference on Computer Vision and Pattern Recognition (CVPR).

[9] Li, H., Lin, Z., Shen, X., Brandt, J. and Hua, G., 2015," *A Convolutional Neural Network Cascade for Face Detection*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 5325-5334.

[10] Bell, S., Upchurch, P., Snavely, N. and Bala, K., 2015," *Material Recognition in the Wild with the Materials in Context Database*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR),3479-3487.

[11] Bertasius, G., Shi, J. and Toressani, L., 2015," *DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection"*, was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 4380-4389.

[12] ITQON, Shunichi, K. and Satoru, I., "*Improving Performance of k-Nearest Neighbor Classifier by Test Features*", was published in the Springer Transactions of the Institute of Electronics, Information and Communication Engineers 2001.

[13]Suguna, N., and Thanushkodi, K. "*An Improved k-Nearest Neighbor Classification Using Genetic Algorithm*" was published in the Conference on Computer Vision and Pattern Recognition (CVPR).

[14]Parvin. H., Alizadeh, H. and Minaei-Bidgoli. B, "*KNN: Modified K-Nearest Neighbor*", was published at the Proceedings of the World Congress on Engineering and Computer Science 2008.

[15]Chen, Q., Huang, J., Feris, R., Brown, L. M., Dong, J. and Yan, S., 2015," *Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 5315-5324.

[16]Chen, X. and Zitnick, C. L., 2015*," Mind's Eye: A Recurrent Visual Representation for Image Caption Generation*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2422-2431.

[17]Krizhevsky, A., Sutskever, I., and Hinton, G. E. "*Imagenet classification with deep convolutional neural networks*". In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[18]Cimpoi, M., Maji, S. and Vedaldi, A., 2015," *Deep Filter Banks for Texture Recognition and Segmentation*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3828-3836.

[19]Dai, J., He, K., and Sun, J., 2015, " *Convolutional Feature Masking for Joint Object and Stuff Segmentation*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3992-4000.

[20]Donahue, J., Hendricks, L. A., Guadarrama, S. and Rohrbach, M., 2015," *Long-term Recurrent Convolutional Networks for Visual Recognition and Description*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2625-2634.

[21]Dosovitskiy, A., Springenberg, J. T. Brox, T., 2015,"*Learning To Generate Chairs with Convolutional Neural Network*," was published in the Conference on Computer Vision and Pattern Recognition (CVPR),1538-1546.

[22]Escorcia, V., Niebles, J. C. and Ghanem, B.,2015," *On the Relationship between Visual Attributes and Convolutional Networks*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR),1256-1264.

[23]Fan, X., Zheng, K., Lin, Y. and Wang, S., 2015," *Combining Local Appearance and Holistic View: Dual-Source Deep Neural Networks for Human Pose Estimation*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 1347-1355.

[24]Fang, H. *et al*., 2015," *From Captions to Visual Concepts and Back*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 1473-1482.

[25]Gan, *et al.*, 2015," *DevNet: A Deep Event Network for Multimedia Event Detection and Evidence Recounting*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2568-2577.

[26]Girshick, R., Iandola, F., Darrell, T., Malik, J., 2015," *Deformable Part Models are Convolutional Neural Networks*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 437-446.

[27]Gkioxari, G. and Malik, J., 2015," *Finding Action Tubes*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 759-768.

[28]Gonzalez-Garcia, A., Vezhnevets, A. and Ferrari,V.,2015," *An Active Search Strategy for Efficient Object Class Detection*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3022-3021.

[29]Han, X., Leung, T., Jia, Y., Sukthankar, R. and Berg, A. C., 2015," *MatchNet: Unifying Feature and Metric Learning for Patch-Based Matching*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3279-3286.

[30]Hariharan, B, Arbel´aez, P., Girshick, R. and Malik, J.,2015," *Hypercolumns for Object Segmentation and Fine-grained Localization*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 447-456.

[31]Hoffman, J., Pathak, D., Darrell, T. and Saenko, K., 2015," *Detector Discovery in the Wild: Joint Multiple Instance and Representation Learning*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 2883-2891.

[32]Hosang, J., Rodrigo, M.O. and Schiele, B. B., 2015," *Taking a Deeper Look at Pedestrians*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR),4073-4082.

[33]Hu, J., Lu, J. and Tan, Y. P., 2015," *Deep Transfer Metric Learning*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 325-333.

[34]Karpathy, A. and Fei-Fei, L., 2015," *Deep Visual-Semantic Alignments for Generating Image Descriptions*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3128-3137.

[35]Klein, B., Wolf, L. and Afek , Y., 2015," *A Dynamic Convolutional Layer for Short RangeWeather Prediction*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 4840-4848.

[36]Lai, H., Pan, Y., Liu, Y. and Yan, S., 2015," *Simultaneous Feature Learning and Hash Coding with Deep Neural Networks*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3270-3278.

[37]Lenc, K., and Vedaldi, A., 2015," *Understanding image representations by measuring their equivariance and equivalence*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 991-999.

[38]Li, Y, Liu, L.,Shen, C. and Hengel, A. V. D., 2015," *Mid-level Deep Pattern Mining*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR),971-980.

[39]Li, S. *et al*., 2015," *Shape Driven Kernel Adaptation in Convolutional Neural Network for Robust Facial Trait Recognition*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 222-230.

[40]Li, G., and Yu, Y., 2015," *Visual Saliency Based on Multiscale Deep Features*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 5455-5463.

[41]Liang, M., and Hu, X., 2015," *Recurrent Convolutional Neural Network for Object Recognition*", was published in the Conference on Computer Vision and Pattern Recognition (CVPR), 3367-3375.

[42]Lin, D.,Shen, X., Lu, C. and Jia, J., 2015," *Deep LAC: Deep Localization, Alignment and Classification for Fine-grained Recognition*", was published in the Conference on Computer Vision and Pattern Recognition,1666-1674.

[43]Lin, T. Y., Cui, Y., Belongie, S. and Hays, J., 2015," *Learning Deep Representations for Ground-to-Aerial Geolocalization*" was published in the Conference on Computer Vision and Pattern Recognition, 5007-5015.