Implementation Automatic Image Tag Assignment and Refinement using Machine Learning Technique

#1Sayali Nadgauda, #2Swati Borude, #3Shubhangi Barhate, #4Ghodake Pranita, Prof.Rathod V.U
#1234Department of Computer Engineering,
Shree Chhatrapati Shivaji Maharaj Collage of Engineering Nepti.

Abstract: **Tag-based image search is one of the important methods to find images contributed by social users in such social websites. How to make the top ranked result relevant and with diversity is challenging Tag-based image search. It is commonly used in social media than content-based image retrieval and context and content-based image retrieval. Social image tag refinement is to remove the noisy or irrelevant tags and add the relevant tags. The testing data is for image tag assignment and images are randomly chosen as the learning data while the rest ones are used as the testing data.**

Keywords: **CNN, image tagging, , tag-based image retrieval.**

I. INTRODUCTION

Machine Learning is an idea to learn from examples and experience, without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves [1]. Image recognition, in the context of machine vision, is the ability of software to identify objects, places, people, writing and actions in images [2]. Deep learning is a part of machine learning algorithms that are recently introduced to solve complex, high-level abstract and heterogeneous datasets, especially image and audio data. There are several types of deep learning architectures, which are deep neural network (DNN), convolutional Neural Network (CNN), deep belief networks (DBN) and convolutional deep belief networks (CDBN)[3]. In real-world applications, many photo sharing websites, such as Flickr and Facebook, have been becoming popular, which facilitate millions of users to upload, share and tag their images. It leads to the dramatic increase in the number of images associated with user-provided tags available for example, it is reported in March 2013 that Flick had more than 3.5 million new images uploaded daily [4]. It sheds new light on the problem of image understanding. Unfortunately, these tags are provided by amateur users and are imperfect, i.e., they are often incomplete or inaccurate in describing the visual content of images, which brings challenges to the tasks of image understanding such as tag-based image retrieval [4].

In this work, we focus on refining image tags to complement relevant tags and remove the irrelevant tags, and assigning tags to new images. Image annotation is traditionally treated as a machine learning problem, which always depends on a small-scale manually-labeled data. However, they fail to handle large-scale social images due to the weakly-supervised data. Different from the traditional image annotation, tag refinement is to remove irrelevant tags from the initial tags associated with images. With the advent of mobile and communication technologies, smart phones and other image capturing applications are increasing day by day. Social media has affected in our daily lives. People are increasingly becoming more interested in posting their daily experience online and sharing their feelings with others Flickr is one of the decent photo sharing websites which contains more than 10 billion photographs from people in different situations. A picture provides wealth information about users' preference, insight and sentiment. This information could be widely used in several fields such as campaign prediction, stock price forecast and advertisement recommendation.

However, these pictures may consist of irrelevant information or sometimes unclear points. Therefore, based on this messy information, it is hard to identify feelings and correct concepts in the pictures.

II. LITERATURE SURVEY

In the multimedia and data mining communities, many researchers focus on the problem of social image analysis. Different traditional image annotation method that usually learn models from small-scale manually-labeled images, these methods exploit massive images associated with weakly-supervised user-provided tags. In this subsection, we present the related work about social image tag refinement and social image tag assignment. Social image tag refinement is to remove the noisy or irrelevant tags and add the relevant tags. In, the group information of images from Flickr is exploited with the assumption that the images within a batch are likely to have a common style. Zhu et al. proposed to decompose the image-tag matrix into a low rank 1matrix and a sparse matrix and considered the content consistency and tag correlation as regularization terms. It can naturally embed new images into the subspace using the learned deep architecture. Besides, to remove the noisy or redundant visual features, a sparse model is imposed on the transformation matrix of the first layer in the deep architecture. Finally, a unified optimization problem with a well-defined objective function is developed to formulate the proposed problem. Extensive experiments on real-world social image databases are conducted on the tasks of image tag refinement and assignment. Encouraging results are achieved with comparison to the state-of-the-art algorithms, which demonstrates the effectiveness of the proposed method. It can naturally embed new images into the subspace using the learned deep architecture. Besides, to remove the noisy or redundant visual features, a sparse model is imposed on the transformation matrix of the first layer in the deep architecture [1].

The number of images associated with weakly supervised user-provided tags has increased dramatically in recent years. User-provided tags are incomplete, subjective and noisy. In this work, we focus on the problem of social image understanding, i.e., tag refinement, tag assignment and image retrieval. Different from previous work, we propose a novel Weakly-supervised Deep Matrix Factorization (WDMF) algorithm, which uncovers the latent image representations and tag representations embedded in the latent subspace by collaboratively exploring the weakly-supervised tagging information, the visual structure and the semantic structure. Due to the well-known semantic gap, the hidden representations of images are learned by a hierarchical model, which are progressively transformed from the visual feature space. It can naturally embed new images into the subspace using the learned deep architecture [2].

Text sentiment analysis has gained a great value in social networks due to its popularity and simplicity. Image sentiment analysis has also attracted a lot of attention through recent years. It is apparent that these approaches, neither text sentiment nor image sentiment analyzes are by themselves sufficient to obtain an accurate performance. On the other hand, the combination of them has compounded the problem. Thus, this paper provides a way to utilize the strengths of these techniques to develop a sophisticated method, called Supervised Collective Matrix Factorization (SCMF). The visual feature and textual feature are represented by Alexnet deep learning network and Bag of Glove Vector (BoGV) respectively. The proposed approach takes label information into consideration during matrix factorization, which is inspired by the graph Laplacian work [3].

Semi-Non-negative Matrix Factorization is a technique that learns a low-dimensional representation of a dataset that lends itself to a clustering interpretation. It is possible that the mapping between this new representation and our original data matrix contains rather complex hierarchical information with implicit lower-level hidden attributes, that classical one level clustering methodologies cannot interpret. In this work we propose a novel model, Deep Semi-NMF,that is able to learn such hidden representations that allow themselves to an interpretation of clustering according to different, unknown attributes of a given dataset. Mainly focus on extensions including sparse, kernel-based, convolutive and a novel supervised dictionary learning variant of principal component analysis and non-negative matrix factorization. We also present a semi- supervised version of the algorithm, named Deep WSF, that allows the use of (partial) prior information for each of the known attributes of a dataset, that allows the model to be used on datasets with mixed attribute knowledge [4].

The similarity between all research papers is that they all are using CNN methods for predicting tags of images. The reason behind using CNN is its accuracy of predicting result.

III. PROPOSED SYSTEM

A. System architecture

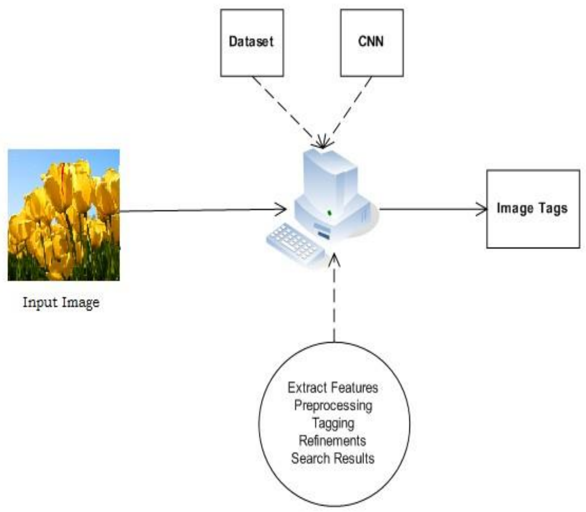

Figure No. 3.1 System Architecture

In the diagram, there is flow of our project.

1. The whole architecture is made by PyQT library used in python language. PyQT library gives all the necessary stuff related to GUI design. PyQT provides us display screen, buttons and so on. So, in this way PyQT helps us in design GUI.

2. After designing of GUI, another task is to authenticate valid user for operating application. To deal with this task, we are using MySQL database to store data of username and password and through this, user can authenticate easily.

3. Another task is to preprocess the input image which can be done by OpenCV library of python. By using this library, image is converted into grayscale image, contour image and smoothen image.

4. The major task of this survey paper is to collect datasets of various images and to achieve this result, we are working on MIR Flicker25k image dataset. Using this dataset, we train a model for assigning appropriate tags to the given input image. If any image has user defined irrelevant tags then system can refine tags.

In this way, we achieve our all the tasks to achieve our project goal.

B. Mathematical expression

Mathematical Model Let S be the Closed system defined as, S = {Ip, Op, A, Ss, Su, Fi} Where, Ip=Set of Input, Op=Set of Output, Su= Success State, Fi= Failure State and A= Set of actions, Ss= Set of user's states.

- Set of input=Ip= {username, password, input image, dataset, CNN data}
- Set of actions =A= {F1, F2, F3, F4, F5, F6} Where, o
  - F1= Authentication of user o
  - F2 = Image preprocessing o
  - F3 = Dataset Preprocessing o
  - F4 = Labeling/ Tagging o
  - F5= Refinement o
  - F6= Searching Relative Images
- Set of user's states=Ss= {login state, input image, view tags, view search results}
- Set of output=Op= {Image Tags, Search results}
- Su=Success state= {Login Success, CNN data process, Image tags prediction, Search Results}
- Fi=Failure State= {Invalid image, Login failed, Dataset read failure}
- Set of Exceptions= Ex = {Null Pointer Exception, Null Values Exception, CNN Exception}

IV. RESULT AND DISCUSION

Fig 4.1. Login page

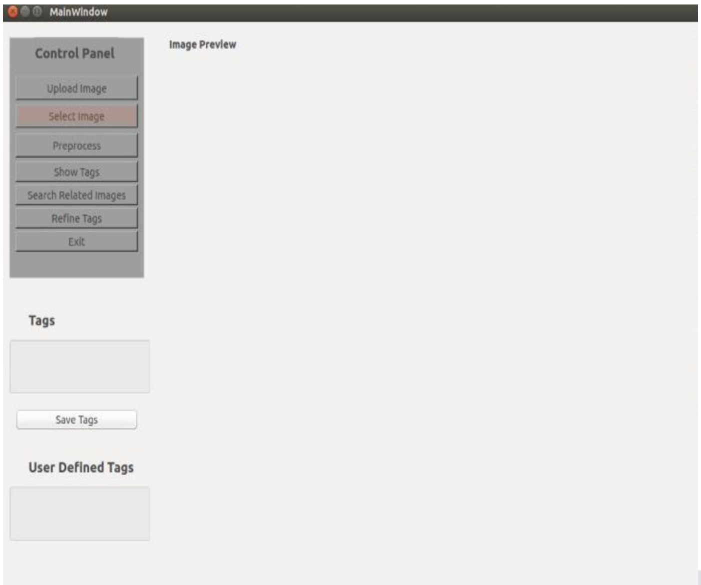

Fig 4.2. Home page

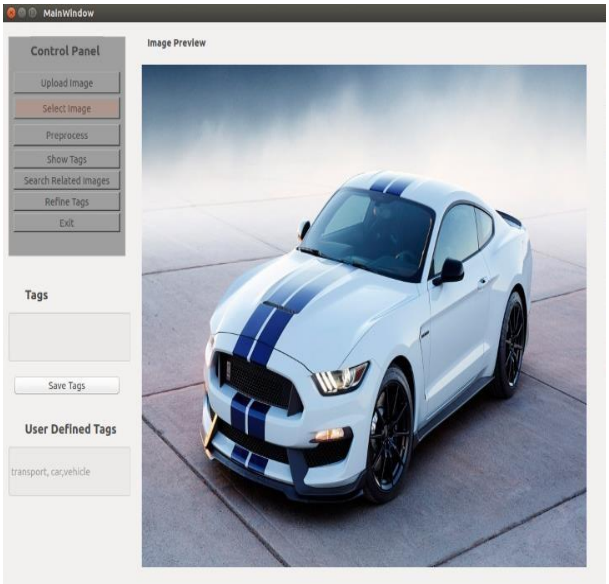

Fig 4.3. Upload Image Preview

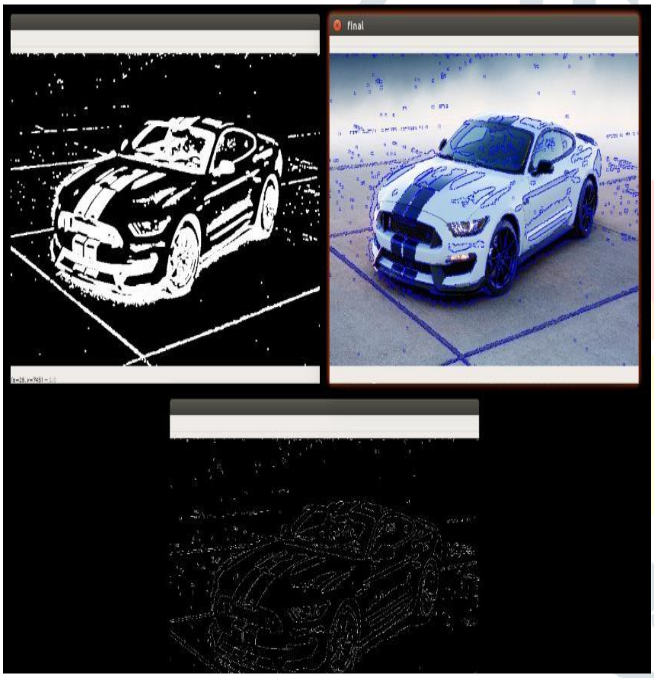

Fig 4.4. Image Processing

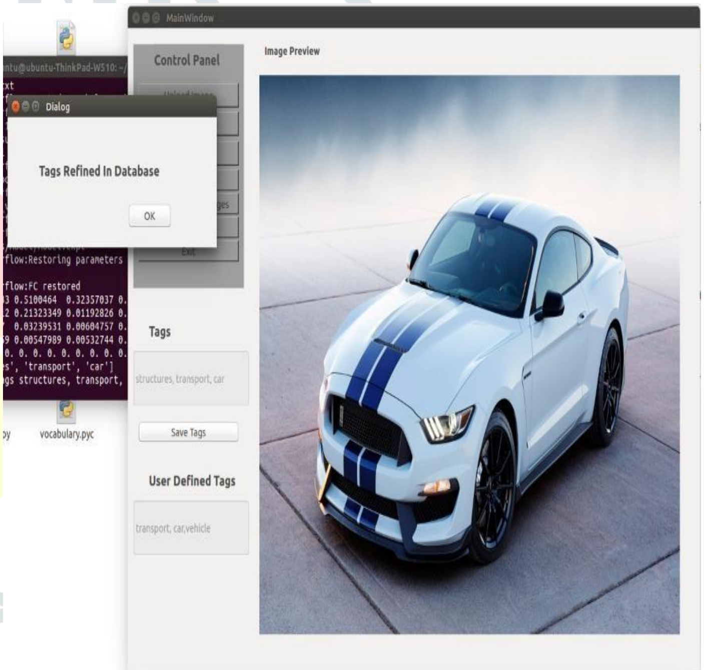

Fig 4.5. Tag generate and stored database

V. CONCLUSION

We propose a weakly supervised convolutional neural network for social image tag refinement and tag assignment method via the deep non negative low-rank model. The visual features and the high-level tags are connected by the deep architecture. The tag refinement and the learning of parameters are jointly implemented, which makes the proposed method have good scalability. Extensive experiments are conducted on two widely used datasets and the experimental results show the advantages of the proposed method for tag refinement and assignment. To well handle the out-of-sample problem, the underlying image representations are assumed to be progressively transformed from the visual feature space.

VI. ACKNOWLEDGMENT

I wish to express my profound thanks to all who helped us directly or indirectly in making this paper. Finally, I wish to thank to all our friends and well- wishers who supported us in completing this paper successfully I am heartily thankful to my project guide Prof. Rathod V. U for his valuable guidance and inspiration. In spite of their busy schedules they devoted their self and took keen and personal interest in giving us constant encouragement and timely suggestion. Without the full support and cheerful encouragement of my guide, the paper would not have been completed on time.

VII. FUTURE WORK

The proposed approach can deal with the noisy, incomplete or subjective tags and the noisy or redundant visual features. In future, we will focus on uncovering the latent structures of data and incorporating it into the proposed model in this work. How to extract representations from raw pixels based on the proposed model is also our future work.

VIII. REFERENCES

[1] https://extertsystem.com/machine-learning definition

[2] https://searchenterpriseai.techtarget.com/definition/image-recognition//

[3] Wang, T., W, D. J., Coates, A., & Ng, A. Y. (2012). End-to-end text recognition with convolutional neural networks. ICPR. Dieleman, S., mon Brakel, P., & Schrauwen, B. (2011). Audio-based music classification with a pretrained convolutional network. ISMIR.

[4] Le, D., & Provost, E. M. (2013). Emotion recognition from spontaneous speech using hidden markov models with deep belief networks.

[5] Hensman, P., & Masko, D. (2015). The impact of imbalanced training data for convolutional neural networks (Unpublished doctoral dissertation). KTH Royal Institute of Technology.

[6] G. Zhu, S. Yan and Y.Ma, "Image tag refinement towards low rank, content tag prior and error sparsity," in Proceedings of ACM Int'i conf. on multimedia ,2010, pp. 461-470.

[7] Zechao Li, Jinhui Tang "Deep Matrix Factorization for Social Image Tag Refinement and Assignment":

[8] Zechao Li, Jinhui Tang, "Weakly Supervised Deep Matrix Factorization for Social Image Understanding":

[9] Liu, Y., Yu, X., Huang, J. X., & An, A. (2010). Combining integrated sampling with SVM ensembles for learning from imbalanced datasets. Elsevier Ltd.

[10] www.google.com/search?q=medium.freecodecamp+cnn + diagram.