Image Tagging using Machine Learning Technique

Mrs. Mahananda R. Tidke, Dr. Shyam Gupta
Department of Computer Engineering,
Siddhant college of Engineering – [SCOE] Sudumbare, Pune, India.

Abstract: **Tag-based image search is one of the important method 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, tag refinement.

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 website 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

An automatic approach to locate relevant image patches and model image tagging within the Multiple Instance Learning (MIL) framework is proposed in this study. The first end-to-end trainable deep MIL framework for the multi-label zero-shot tagging problem is proposed. It explores several alternatives for instance level evidence aggregation and performs an extensive ablation study to identify the optimal pooling strategy. Due to its novel design, the proposed framework has several interesting features: (1) unlike previous deep MIL models, it does not use any offline procedure (e.g., Selective Search or EdgeBoxes) for bag generation. (2) During test time, it can process any number of unseen labels given their semantic embedding vectors. (3) Using only image-level seen labels as weak annotation, it can produce a localized bounding box for each predicted label. We experiment with the large-scale NUS-WIDE and MS-COCO datasets and achieve superior performance across conventional, zero-shot and generalized zero-shot tagging tasks.[1]

In this paper, Author proposes a new way of measuring tag preferences, and also proposes a new personalized tagging objective function that explicitly considers a user's preferred tag orderings. We also provide a (partially) greedy algorithm that produces good solutions to our new objective and under certain conditions produces an optimal solution. We validate our method on a subset of Flickr images that spans 5000 users, over 5200 tags, and over 90,000 images. Our experiments show that exploiting personalized tag orders improves the average performance of state-of-art approaches both on per-image and per-user bases.[2]

The main contribution of this paper is derivation of a novel framework to refine visual features of tagged images based on graph trilateral filter-based smoothing. This enables reduction of the influence of noisy tags that are irrelevant to contents of images. As a result, accurate BLL becomes feasible by nearest neighbor search using the refined visual features.[3]

This paper introduces a new development approach of intelligent tag system. The designed system can be used to distribute video resource classification and content retrieval method. The system can comprehensively combine video audio, image and

subtitles to get video information. On this basis, it can output more precise tags and a wider range of semantic tags through iterative algorithms. The results of the simulation demonstrate that the tags from the system which is supported by a real-time updated database are significant. By comparing with other existing solutions, the designed system in the paper has better performance from valid tag quantity, valid tag rate and Dispersed Contribution Coefficient.[4]

In this study, the main aim to refine image captions by utilizing Self Organizing Maps. In this the main part is to extract image and caption pairs as feature vectors and then cluster those vectors. Vectors with similar content clustered close to each other. With the help of those clusters, it hopes to get some relevant tags that do not exist in the original tags. Author performed extensive experiments and presented in the initial results. According to these results, the proposed model performs reasonably well with a 54% precision score.[5]

In this paper, Seq-CVAE is proposed which learns a latent space for every word. In this temporal latent space to capture the 'intention' about how to complete the sentence by mimicking a representation this summarizes the future. The efficacy of the proposed approach on the challenging MSCOCO dataset, significantly improving diversity metrics compared to baselines while performing on par with respect to sentence quality.[6]

In this paper, Author formulate the problem of semantic image hashing as a weakly-supervised learning problem. Author utilizes the information contained in the user-generated tags associated with the images to learn the hash codes. More specifically, It extract the word2vec semantic embeddings of the tags and use the information contained in them for constraining the learning. Accordingly, Author names their model Weakly Supervised Deep Hashing using Tag Embeddings (WDHT). WDHT is tested for the task of semantic image retrieval and is compared against several state- of-art models. Results show that our approach sets a new state- of-art in the area of weekly supervised image hashing.[7]

In the transmission and data processing communities, several researchers concentrate on the matter of social image analysis. Completely different ancient image annotation ways that sometimes learn models from small-scale manually- labeled pictures, these ways exploit large pictures related to weakly-supervised user-provided tags. During this section, we tend to gift the connected work regarding social image tag refinement and social image tag assignment. Social image tag refinement is to get rid of the rip-roaring or digressive tags and add the relevant tags. In, the cluster data of pictures from Flickr is exploited with the belief that the pictures inside a batch area unit probably to possess a typical vogue. It will naturally enter new pictures into the mathematical space victimisation the learned deep design. Besides, to get rid of the rip-roaring or redundant visual options, a thin model is obligatory on the transformation matrix of the primary layer within the deep design. Finally, a unified improvement drawback with a well-defined objective perform is developed to formulate the projected drawback. in depth experiments on real-world social image databases area unit conducted on the tasks of image tag refinement and assignment. Encouraging results area unit achieved with comparison to the progressive algorithms, that demonstrates the effectiveness of the projected methodology. It will naturally enter new pictures into the mathematical space victimisation the learned deep design. Besides, to get rid of the rip-roaring or redundant visual options, a thin model is obligatory on the transformation matrix of the primary layer within the deep architecture[8].

The number of pictures related to feeble supervised user-provided tags has exaggerated dramatically in recent years. User-provided tags area unit incomplete, subjective and rip-roaring. during this work, we tend to concentrate on the matter of social image understanding, i.e., tag refinement, tag assignment and image retrieval. completely different from previous work, we tend to propose a completely unique Weakly-supervised Deep Matrix factoring (WDMF) rule, that uncovers the latent image representations and tag representations embedded within the latent mathematical space by collaboratively exploring the weakly-supervised tagging data, the visual structure and therefore the linguistics structure. because of the well-known linguistics gap, the hidden representations of pictures area unit learned by a gradable model, that area unit increasingly reworked from the visual feature area. It will naturally enter new pictures into the mathematical space victimisation the learned deep design [9].

The similarity between all analysis papers is that all of them area unit victimisation CNN ways for predicting tags of pictures. the rationale behind victimisation CNN is its accuracy of predicting result.

.

III. PROPOSED METHODOLOGY

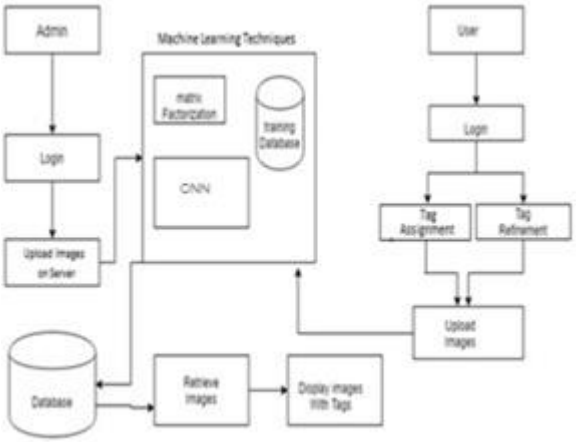

Fig 1. System architecture

The architecture of the proposed system is as shown in fig.1.

In this figure, there are two login models shown to access the system. First model is for Admin login and another for User login. Admin can upload images on the server. User should upload images to the system and according to the uploaded image; tag

should be shown to the user. Tag should be shown after the machine learning process.

We are using MirFlikr image dataset which consist of 25000 images along with annotations and json data. This dataset represents 24 annotations or 24 words/ tags. We used CNN to train this dataset and get weights in model file for each annotation. This weighted model further can be used to predict 24 tags from dataset on given image.

As the dataset is large and its size is around 3.6 GB we will need high GPU computing to train such dataset. We used Crestle.io web platform to train this dataset along with GPU computing.

A. Algorithm:

CNN algorithm is used to train the dataset as follows,

1. Read the dataset and annotations
2. Generate the Json file for annotations as well as image data.
3. Read features of all images and label (here annotations will act as labels) of it using following functions,

a. Conv2D:

Read dataset and convolute through epochs for forming a network layer

b. Maxpool2D:

Fetch maximum features from images

c. Relu activation for layers:

Plot linear network layer by rectifying data points

d. Sigmoid activation for dense layer:

Non-linear connections in dense layer are optimized and form fully connected network layer

e. Categorical Crossentropy for loss calculation: Calculations of loss and evaluating model based on metrics accuracy.

4. Store it in model file
5. Get input image
6. Read features of input image
7. Compare features of stored features
8. Show label as prediction of nearly matched features.

B. 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,
  - F1= Authentication of user
  - F2 =Image preprocessing
  - F3 = Dataset Preprocessing
  - F4 =Labeling/ Tagging
  - F5= Refinement
  - 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 ={NullPointerException, NullValues Exception, CNN Exception }

IV. RESULT AND DISCUSION

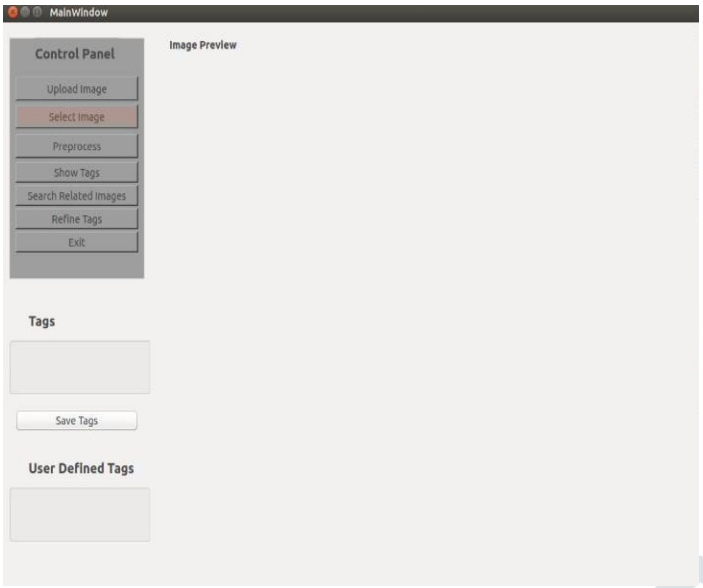

Fig 4.2. Home page

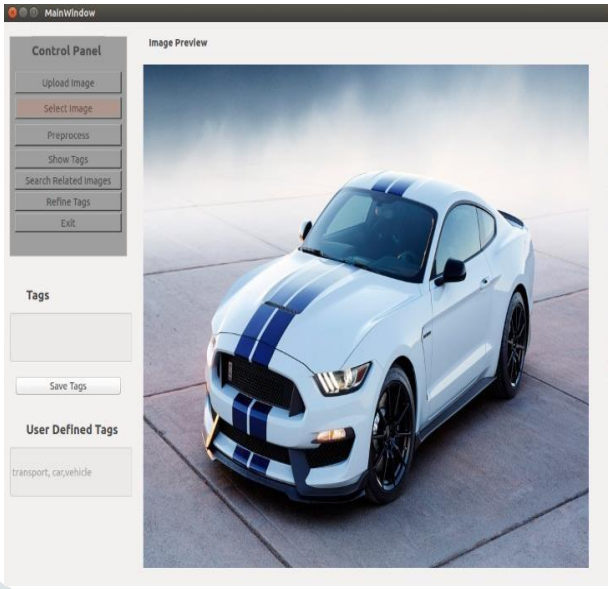

Fig 4.3. Upload Image Preview

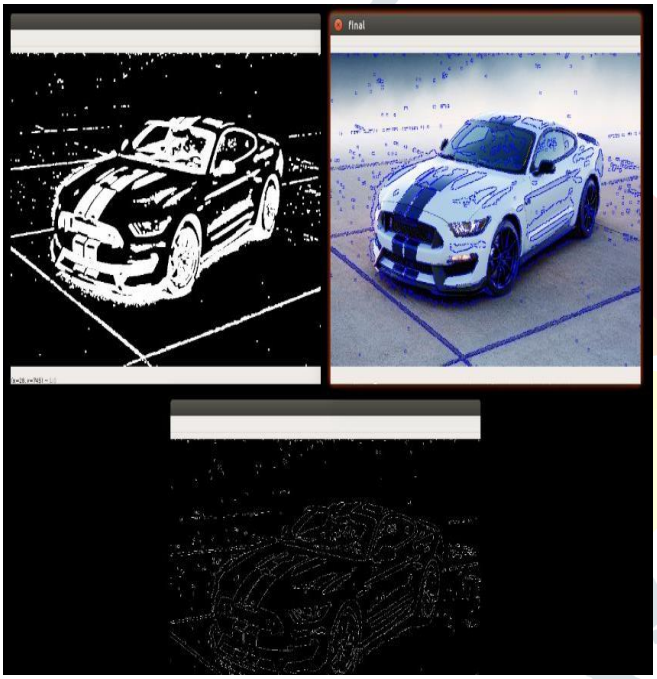

Fig 4.4. Image Processing

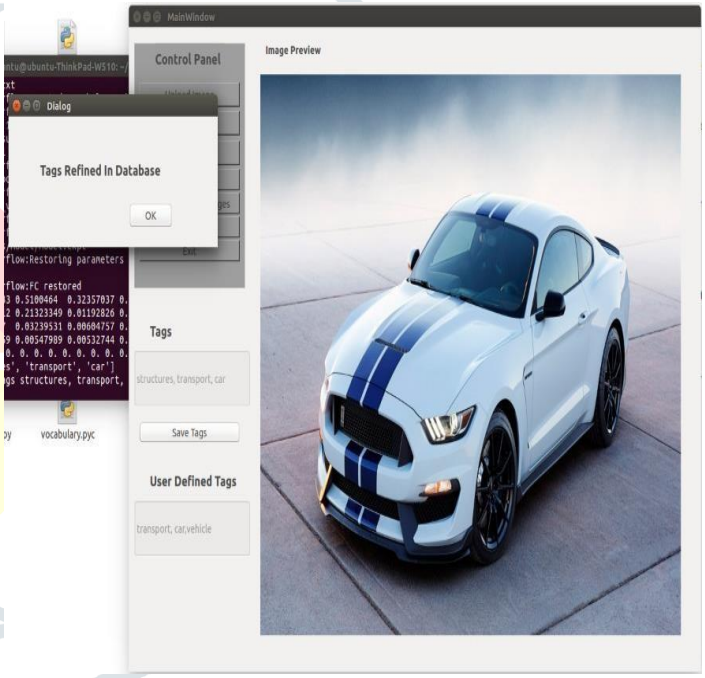

Fig 4.5. Tag generate and stored database

In the proposed system, we will be using standard dataset along with the annotations so that we can have more accuracy. The dataset used is Mirflikr dataset with 24 annotations and 25000 images which is large enough to give accurate predictions on 24 tags. The size of dataset is larged and it is trained by using GPU computing so that model will be more accurate.

Comparative results of existing and proposed system is as follow,

Parameters	**Existing System**	**Proposed System**
Mirflikr Dataset	Somewhat	Yes
CNN	No	Yes
GPU Computing	No	Yes
Annotations	Somewhat	Yes
Json Data	No	Yes
Fast prediction	No	Yes
Portable Model	No	Yes

Table 1: Comparative Results

With reference to Table 1 it is clear that we overcome various problems in existing system and our approach works efficiently.

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.
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.

VI. REFERENCES

[1] Shafin Rahman, "Deep0Tag: Deep Multiple Instance Learning for Zero-shot Image Tagging", 1520-9210 (c) 2019 IEEE.

[2] Amandianeze O. Nwana Tsuhan Chen," Who Ordered This?: Exploiting Implicit User Tag Order Preferences For Personalized Image Tagging", 2017.

[3] YuiMatsumoto ; Shota Hamano ; Ryosuke Harakawa ; Takahiro Ogawa ; Miki Haseyama, "Bilingual Lexicon Learning Using Tagged Images via Graph Trilateral Filter-based Feature Refinement", May 2019,IEEE.

[4] Shan Liu ; Fengxuan Shao, "Development of Intelligent Tag System", 10.1109/CISP-BMEI48845.2019.8965966, Oct. 2019 IEEE.

[5] Tolga Üstünkök ; Ozan Can Acar ; Murat Karakaya, "Image Tag Refinement with Self Organizing Maps", 10.1109/UBMYK48245.2019.8965477, Nov. 2019 IEEE.

[6] Jyoti Aneja ; Harsh Agrawal ; Dhruv Batra ; Alexander Schwing," Sequential Latent Spaces for Modeling the Intention During Diverse Image Captioning", 10.1109/ICCV.2019.00436, Nov. 2019 IEEE.

[7] Vijetha Gattupalli ; Yaoxin Zhuo ; Baoxin Li, "Weakly Supervised Deep Image Hashing Through Tag Embeddings", 10.1109/CVPR.2019.01062, June 2019,IEEE

[8] XuemingQian , Member, IEEE, Dan Lu, and Xiaoxiao Liu, Tag Based Image Search by Social Re-ranking,12 May 2016.

[9] VilasDilipMane,Prof.NileshP.Sable,"TagBasedImageSearchb ySocialReranking,12 December 2016.

[10] Shweta gonde1, Assistant Professor Uday Chouarisia2, Assistant Professor Raju Barskar3,"A SURVEY ON WEB IMAGE SEARCH USING RERANKING 5 may 2014.

[11]A.Ksibi,AB.Ammar,CB.Amar."Adaptivediversification for tag-basedsocialimage retrieval". International Journal of Multimedia Information Retrieval, 2014, 3.1: 29-39.

[12]V Rajakumar, Vipeen V Bopche, Image Search Re-ranking ,5 Dec. 2013.

[13]X.Qian,H.Wang,G.Liu,X.Hou,"HWVP:HierarchicalWavelet PacketTexture Descriptors and Their Applications in Scene Categorization and Semantic Concept Retrieval", Multimedia Tools and Applications, May 2012.

[14]Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao. "Tag-based social image search with visual-text joint hypergraph learning". Proceedings of the ACM International Conference on Multimedia information retrieval, 2011:1517- 1520.