“AUTOMATED IMAGE TAGGING AND DETECTION”

Mahak Sharma
mahaksharma3016@gmail.com

Abstract— **Computer Vision has evolved as a great field for Image tagging and manipulating activities.Various researches have been done previously in the domain of Image tagging to differentiate among factors including illuminations, facial features, postures, scaling etc as to achieve improved recognition rates and tagging accuracies .In this paper we’ll go through various research contributions done by authors in the area of Image automation and analyse their strength and weaknesses.The reasons come from the need of automatic recognitions and surveillance systems, the interest in human visual system on face recognition, and the design of human-computer interface and access control. These researches involve knowledge and researchers from disciplines such as machine learning, software developers, neuroscience, psychology, computer vision and pattern recognition. We’ll also analyse several automation tool which help in face and facial features detections through improvised algorithms, colour segmentation techniques and image filtering.**

Keywords—AUT, API,PDF, CSV,KPI,IAGA,NCC,GA, geo-spatial image tagging, RANSAC,PARAFAC

I. INTRODUCTION

Because of the ever increasing demand for technology, Image automation and tagging are recently experiencing a massive shift in popularity.The terminology “Computer Vision” is not easy to describe, just as it’s not easy to explain what it means to “see something” .The baseline model is building some high-level representation, or structure, of images or videos depicting the scenario of real life. The exact nature of this representation depends on the particular application, and consequently “computer vision” is an umbrella like term that covers a wide range of specific problems. Applications belonging to this era describe human behaviour and their body features as well. As the computation power and technology available to us is becoming more profuse, and with freely available libraries such as OpenCV and Tensorflow becoming increasingly mature and trained, we find ourselves living in a world that almost seemed fantasy yesterday: a world where machines not only can drive cars, but also assist visually impaired users in describing the world to them in words.

One of the problems considered in computer vision is classifying images: a predefined set of distinct categories, determining the one that best matches the given image. Another one is detecting objects and patterns or discerning individual objects on the picture; this is often used in labelling or tagging the image with a small number of word tags which are understandable by humans as well as some automated robot and chat bots.The last task is interesting, as a number of websites have sprung up recently which offer exactly this as a service. You upload an image via a RESTful APIs, and you get back a JSON containing the tags. Literature study has been done further to find out how we can explore more in the field of image automation and tagging in the sections mentioned below.

II. PROBLEM DEFINITION

Image tagging is key aspect for image content understanding and image processing based on text patterns.Many computer algorithms requires hundreds or even thousands of human guided training sets and illustrative to learn how to function as the best of all.Image tagging is a form of supervised learning technique which learns to classify on the basis of pictorial data set.For the procedure of image classification image classes constitutes of set of images which are similar (semantically) and therefore can gave at least one common gloss.An image can also be provided with more than one set of common annotations and because of this one image can belong to multiple classes.

The image learning and tagging models are trained on smaller dataset of manually tagged images which are further used to divide images in one or multiple number of classes.All the corresponding training corpuses gives a unique relational mapping between the labelling and the described entities which are semantically similar with the image .Given a novel test data ,the labelled learner compares the visual wordings with an unknown test image, describing an image with some textual wordings in the description box.Many different method has been proposed and discussed in the literature survey to learn and evaluate the dependencies and similarity between the visual constitutes of the image as well as the associated text tag .A similarity trend between the most common approaches to automatically tag an image is that all of them try to predict each subset keyword for an image in an independent fashion (not considering other words for that particular image).Any correlational similarity is taken between the words is also taken into consideration so as to find similar images with a given test image from a corpus of various images which belong to multiple categories.Further in section 3 we have analysed different work done on the Image Automation as well as tools used in corporations and as open repositories for the users to get suitable tags for any image.Section 4 discusses ideal methodologies available from many research papers and experimental works.We have also identified gaps in the studies done and have concluded the result as well in the later sections.

III. RELATED WORK

Various automated image tagging tools are developed nowadays and are being improvised later with the help of learning methodologies. Such tools help in validating complete images including layout and appearances of visuals provided using application under test procedures(AUT) ,by running on different browsing platforms, different screen resolutions many other devices and operating systems. Georgiev et.al[1] developed the next generation API tool Imagga which can also be used as an add-on on your different browsing platforms.It describes what's there in a photograph by returning a list of detected tag sequences or words and it also assigns a confidence score of each corresponding tag.The confidence score is numerical value which represents the confidence level of predicted category.1 signify 100% confidence . Cloudinary Image Upload API is used to upload images into Imagga console and response records include suitable tags for a particular image.Ke performance Indicators(KPIs) identify trends in set of image contents which have strong impact on other set of images. "Dataturcks" developed by Gupta et.al[2] is a full power online tool which includes uploading data which can be zip of images, CSV file with textual data , PDF , Word , Doc or even pages . It provides a user friendly interface for image tagging.Bounding boxes are used to capture the tagged section of any image which are also used to give multiple tags by a user or more.On one set of project multiple number of users can communicate which makes it a open source platform for outsourcing.Users can add multiple images and further add their teammates as contributors and than they can just start tagging images manually which can be used as training dataset for multiple open ended classification projects.

Open ended repository tools such as Sloth by Nilsonholger[3] help in growing era of research in this domain of Computer vision.It is an open-ended tool with high level of flexibility.It allow users to label video files and images . It can easily identify a person from surveillance and also help to know that whether a particular person has appeared in the records more than one time.It is used to label faces, person's bounding boxes in multiple shapes such as rectangular, circles and squares etc, to extract and point out facial features points, locations of body joints, genders , personality traits and many more attributes.Users are able to add multiple number of tags for a particular image or a video captured image .Every tag given to the image makes a Key-Value pair which allows us to perform more detailed file processing. Users can also add-on key types which differentiates point labels from given set of pre-defined labels .For example, labelling for right eye features and left eye features is differentiated by using key value pairs for both the given eyes.

IBM Watson Visual recognition Services[4] uses different set of deep learning algorithms to analyse image data .It also provides us with insights of visual content ,trends with time. We can organise and use various image libraries available by date, understand any individual image and extract out important information , it also recognises food items, create customised classifiers to provide specific set of results that is tailored according to our needs.It is used for diverse applications and industries including Manufacturing, Visual auditing, Social Listening, Insurance ,Social commerce as well as in the retail and education centres.The system firstly, gathers the data and analyse it accordingly.It uses built in capabilities of the service model or you can even customise the service model and the results are shown in the visual content result section of the model. As a service on the cloud system IBM Watson is more secure to use and provide privacy control features for the users.The top. Level categorisation includes Animals(birds, reptiles, mammals, amphibians etc), food(cooked food and beverages), plants(trees, shrubs, vegetables and aquatic plants), people and people oriented information and activities, Sports, Nature, Transportation(Land, Water and Air) and many more including furnishing equipments, fruits, tools, colours, gadgets and devices.The Create a classifier calls require that the user should provide at least two positive examples(.zip files) and one positive or a negative classifier to train and classify model efficiently. There are limitations of file sizes for .zip files of images and a particular image should be 224 X 224 pixel size.Another outstanding tool by Amazon Web Services[5] "Amazon Rekognition" tool make it easy to add video and image analysis to our applications. User can just provide any image or video to the Amazon Rekognition API , and the tool identifies the people , objects, textual data ,scenes, activities as well as detection of any inappropriate content in the image.The tool provides highly accurate facial analysis and facia recognition on different images and video data scripts.We can detect , analyze and compare faces for a wide variety of features such as people counting, user verification and public safety use cases. Key features provided includes Facial Recognition and analysis, Pathing for post game analysis, Unsafe content detection, Celebrity recognition by using learned models and text in image identification etc.It is used for facial recognition , sentiment analysis also provides a searchable library for videos, generates metadata from uploaded videos so user can create a search index for names of celebrities and their time of appearances at a particular place.

LogoGrab by Boschin et.al[6] is an amazing tool for identification of logos .It realised the upcoming brands need to catch up the visual age and trends. They discovered a state of art image detection technology system that allowed brands

1	v is score vector with length |L| (|L| is the number of labels in dataset)
2	Compute Euclidean distance between new image and all training images
3	Select k nearest neighbor $I1, I2, \ldots, Ik$ and compute their weight $W_1, W_2, \ldots, W_k$
4	$v = \sum_{n=1}^{k} w_n * A(I_n, :)$

Fig. 2. KNN Pseud code

that contains their own logos. This type of helpful methodology ensures brand promotion and profitable responses from the audience side. The developed technology is strong and powerful enough to find parts of logo and even detects the misuse of brands. The world leading tools for visuals identification and labelling are coming to the market with more improved version ever other day. Various Machine Learning and Deep learning algorithms such as CNN with gaussian, Salt and pepper noise and Poisson count statistics by Guan et.al[7] are used for identification of images through different obstacles and hurdles.Later section consist of the detailed methodology for the algorithms.

IV. CRITICAL REVIEW OF RESEARCH PAPER

Automated Image Annotation developed by Bahrami et.al[12] challenged the most difficult tasks for image tagging and labelling with a huge number of image dataset.The authors used two well known dataset of Corel5K and IAPR TC- 12. Image Annotation Genetic Algorithm (IAGA) was implemented which consisted of three different phases:In the first phase the high dimensional dataset problem was solved and feature selection was done, set of initial chromosomes were decided. For every chromosome every feature was setup into a binary coded representation. Fitness function was decided , Crossover and mutation was performed. Procedure is continued till the maximum iteration is achieved and the chromosome subset cannot be further optimised , after that procedure is terminated.When the procedure terminates, the fittest individual in the matrix that represents the combined results of classifiers.

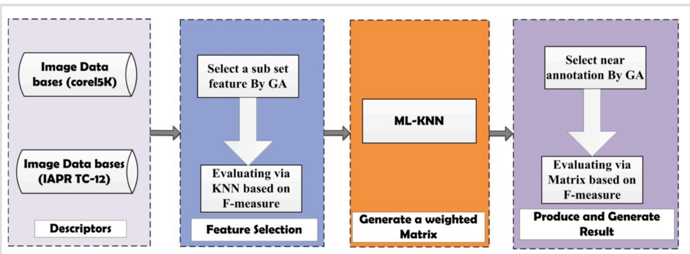


Fig. 1.Proposed IAGA Architecture

After feature selection phase second phase was employed which includes supervised classification technique K-nearest Neighbour classifier(KNN).In this methodology a neighbour with smaller distance is weighted more heavily than one with the greater distance the nearest neighbour gets a weight of 1, the furthest neighbour a weight of 0, and other weights are calculated by the following equation:

$$Wj = dk\text{-}dj/dk\text{-}d1 \qquad \text{if } dk!=d1$$

$$Wj = 1 \qquad \text{if } dk=d1 \qquad (1)$$

In the third phase, combined results are calculated along with weighted matrix of KNN to present final tagging for their test images.Models were evaluated on various model such as Lasso Model.Further models were evaluated and accuracy matrix was setup.

Automated identification of objects based on Normalized Cross-Correlation and Genetic Algorithm by Rusu[13] was a hybrid made up of Normalised Cross-Correlation(NCC) combined with Genetic Algorithm (GA) in order to improve performance of image matching.This Template matching method constitutes of image browsing with a sliding window template to find the maximum point of matching section from the image.Implementation of GA to generate rotations and scaling values was a solution when a particular template differs in aspects of size and angle of rotations for a particular image scene.

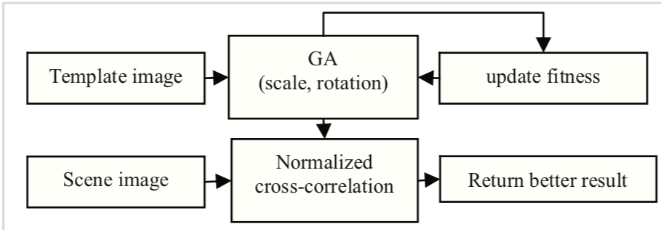


Fig. 3.Block diagram for NCC and GA

In case of Retailing and E-commerce Biased decisions of users can also help in improving existential automated tagging systems. By using NUS-WIDE Dataset Nwana[14] proposed a brand new measurement of tag preferences and also demonstrated that there is indeed a tag-order biasing mechanism.To understand it , when a user mentions tag “a” before tag “b” for a given image, the user is stating that a is of greater importance than b to him.This biased method can be learned from historical time data and further showed that exploitation of biased decisions information can be used to improvised tagging models.We estimate this quantity as the number of times a occurs before b, divided by the number of times they occur together, regardless of order. This allows for anti-symmetry: $p_{ba} = 1 - p_{ab}$.

On the other side, Geo Image tagging is increasingly used for improving image storage, search and retrieval. Sadnes[15] implemented Geo Spatial tagging of images which relies on GPS-enabled cameras, where the geographical location of photographer is stored in the files. The baseline model is based on image timestamp profile only, and the enhanced model improves the accuracy by incorporating the optical cameras for the image settings and compares this with the elevation path track of sun. The strategy was implemented to three image collections where upon the approximated locations, the latitude and longitude of the photographer were determined. The authors outlined a simple and efficient strategy to make intelligent and efficient guesses about the location(a photograph was taken into consideration). Furthermore, exploiting the exposure times and settings of aperture of each and every image, a more precise and accurate longitude estimate was obtained. Another domain of X-ray technique is a key method in modern synchrotron facilities towards material analysis and discovery via characterisation of structures at the molecular as well as nano scale.Meister et.al[7] used the CNN model for training NSLS-|| data for image identification and tagging. The model was analysed with different types of noise including Gaussian, Salt and Pepper Noise and poison counting statistics.5,000 X-ray images were generated and corrupted with three levels of Gaussian Noise, salt and pepper and poison counting stats in this way data augmentation was introduced in training CNN.The best method was with gaussian noise to determine image.

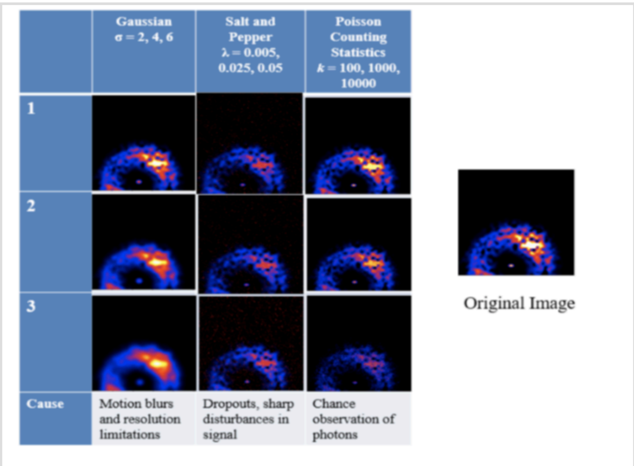
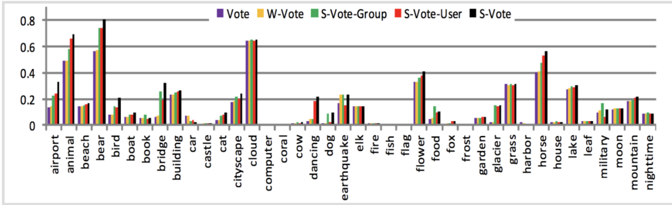


Fig. 4.Noise and image degradation techniques used for data augmentation.

Social tags and audience voting have also been utilised for image searches and developing recommendation systems, however the tags may be biased and noisy. Zhu et.al[8] proposed to extract and leverage image social information, such as labelled preferences of an image owners and social troops that an image has been shared with, by adopting a methodology of well-known neighbour voting approach for automated image tagging .The work proposed a clear picture of socially oriented neighbour method for learning tagged image relevance The sociality is exploited from the image perspective, rather than users on social Websites. The ownership information helps to infer user preferred tags, and the group information provides a strong clue of appearance of the major semantics in target images. The social strength with target images is also considered. Experiments on NUS-WIDE dataset demonstrate the superiority of proposed methods for tag recommendation using either user original tag list or manually justified tags as ground-truth. In addition, general image search can also benefit from the learned relevance using image social clues.

Figure 5 below depicts overall performance of voting of Social clues over different set of queries performed.Sankar et.al[9] invented a new Web image search portal to search and retrieve matched images from the uploaded online images with the help of Auto Hidden tags mentioned using Ransac Algorithm through opened Computer Vision. When a user posting their images, if they forget to comment on all the objects present in the image also their proposed system will detect all the object present in the image using object detection algorithm, then this

Fig. 5.Plot of Voting Queries



object will be extracted from the image and recognised with large dataset by using RANSAC algorithm.This implementation consist of two procedures, one is Offline mode process and another is the online mode process. The Offline process involved activities when the user uploads an image in the Portal and the online process involved when a user is providing search keyword and getting the desired search image results. The main objective of the work was to extract out more relevant images through automated hidden tag approach. In web information retrieval system the image search user to give query keywords as an input, it finds the relevant image as the output. Another algorithm used was the Parallel Factor Analysis 2(PARAFAC2) discovered by Pantraki et.al[10] which is applied to a collection of three matrices, namely the **image-feature matrix**, whose columns constitutes representations capturing the visual appearance of images, the **image-tag matrix**, whose columns indicate the tags associated with each visual image, and the **image-user matrix**, whose columns identify who has uploaded or is associated to each image. PARAFAC2 has been employed for semantically oriented feature extraction of images, multi- label image tagging and annotations, and image recommendation to multiple users. The inclusion of three slices in the PARAFAC2 model enables capturing the latent relations between the images features, the tags, and the user interests. Promising results were reported.Automated image annotation by Sergieh at.al[11] aims at associating unlabelled images with keywords /tags that describe their relevant contents. Early research on automatic annotation techniques focused on using machine learning techniques. The idea was to use a dataset of already labeled images in order to train models for predicting labels for un-annotated images. Designed workflow for annotation model is shown below in Fig.6. which mainly involves data preparation phase(which includes data crawling strategy to collect image data from community photo websites) and a data mining phase.Tag cleaning was done for preprocessing of data and their similarity based on geographies, images and rankings are calculated in the mining phase.

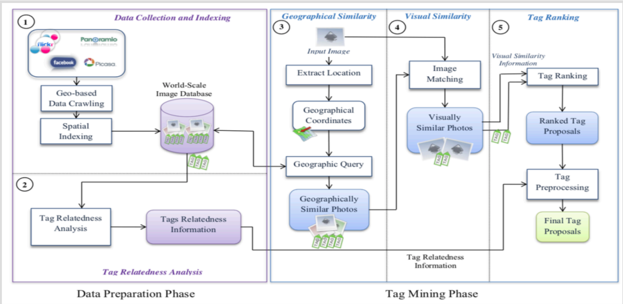


Fig. 6.Workflow for Annotation system

Automated Image tagging based on Contextual information by Evertsen et.al[16] thesis with the University of Tromso is to design, discover and evaluate a system that automatically finds annotations for images based on location (GPS coordinates), date or time and image categories.The database was Flicker shared which is highly unreliable and noisy.The image tagging system was implemented in the thesis can basically be divided into two phases. The first part consists of retrieving a set of images that are considered to be relevant for the selected query image .The second part consists of collecting and processing the tags of the images in the image set found in the first phase of the approach.

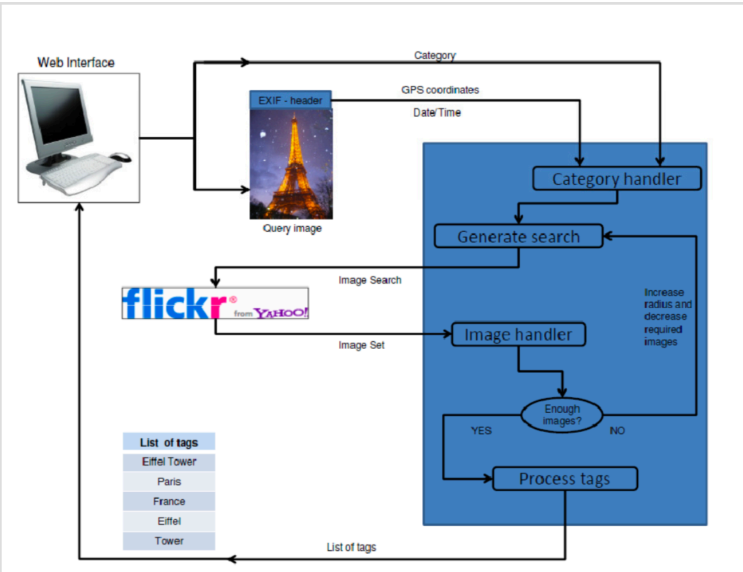


Fig .8. LOCATION tagger System

Tagger system was hard coded in python with set of assumption for the attributes further the system was evaluated ,results are mention in the below section.The runtime of the tagger was around 22 seconds on an average, which was quite faster when many relevant images from different set of users are fouded close to the given test(query) image.

V.CONCLUSION

Obtaining high-quality labeled data is a development barrier that becomes more significant when complex models are developed.With a variety of tagging tools available online in the market, the main challenge for a data science team is to estimate which software will work best for a specific project in terms of functionality and cost. In addition to manual labelling approaches, data scientists have found new methods that partly automate the process and reduce the need for human involvement. I believe the development of these approaches will be the main trend in the near future. Literature stud helped in identifying algorithms like NACC with GA, KNN, CNN, PARAFAC2 as well as Ransac for Automated Image Tagging.Location Tagger system[16] was compared against Simple and Spirit Taggers out of which the model in the research outperformed with an accuracy score of 0.90 along with contextual data. Implementation GA for generating rotation and scaling values in [13]is a better solution when template differs in size and rotation angle of the image scene. If it is used at the distance scale from 0.3 to 3.0 and rotate up to 360 degrees with combining each set of calculations would lead to bulky and would not fit into real time.systems.GA further helps in reducing calculations by reaching to optimal result in reasonable amount of time.IAGA model in[12] was compared with different models and it outperformed on other

Model Name	P	R	Fmeasure	NZR
[13]2007- SML	0.23	0.29	0.26	137
[16]2007- PLSA-WORDS	0.14	0.2	0.16	105
[20]2010- PLSA-FUSION	0.19	0.22	0.2	112
[21]2010- TMIML	0.23	0.27	0.25	130
[15]2010- JEC	0.27	0.32	0.29	139
[15]2010- Lasso	0.24	0.29	0.26	127
[22]2011-GMM	0.15	0.19	0.17	93
[23]2011- GM-PLSA	0.25	0.26	0.25	125
[24]2011- CKSM	0.29	0.35	0.31	147
[25]2012- HDGM	0.29	0.3	0.3	146
Proposed method - IAGA	**0.3**	**0.327**	**0.31**	**132**

models on COREL-5K dataset

Fig. 9.Results for IAGA and other model on COREL-5K data

Parafac2[10],In order to evaluate the proposed tagging method, the dataset was randomly split into a training and test set at a ratio 60% and 40%, respectively.With 1264 images, 512 features, 2366 tags and 440 users along with the k value 20 results of parafac came out to be better as compared to Random forest and upper bounded models as shown in Fig.10.

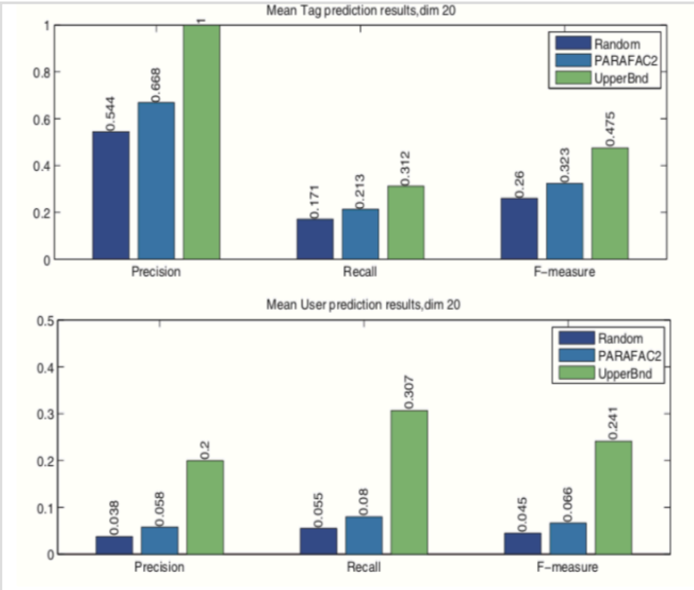


Fig. 10.Comparison of PARAFAC2 with other set of Models

VII.FUTURE WORK

We believe that there are several methods that the work in the era of Image tagging could be extended or extend through other works. One way is to work upon algorithms and methods that provide strong guarantees for solving our tagging objective of images , or even solves it optimally, even in the presence of dependencies and relational model. Another direction comes from the observation that most tag pairs never occur at the same time, but it may be possible to learn a function that maps a pair of tags to a real number indicating the direction and strength of the preferences.Another methodology is to combine meta-information with image contents to achieve more reliable and efficient accuracy results.

VIII.REFERENCES

[1].https://imagga.com/.

[2].https://dataturks.com/

[3].https://github.com/cvhciKIT/sloth

[4].https://www.ibm.com/watson/services/visual-recognition/

[5].https://aws.amazon.com/rekognition/

[7].Meister, Nicole, et al. "Robust and scalable deep learning for X-ray synchrotron image analysis." *Scientific Data Summit (NYSDS), 2017 New York*. IEEE, 2017.

[8].Zhu, Shiai, Samah Aloufi, and Abdulmotaleb El Saddik. "Utilizing image social clues for automated image tagging." *Multimedia and Expo (ICME), 2015 IEEE International Conference on*. IEEE, 2015.

[9].Sankar, K., and GNK Suresh Babu. "Web information retrievals: An excellent image portal with automated hidden tag to image." *Advances in Computer Applications (ICACA), IEEE International Conference on*. IEEE, 2016.

[10].Pantraki, Evangelia, and Constantine Kotropoulos. "Automatic image tagging and recommendation via PARAFAC2." *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*. IEEE, 2015.

[11].Sergieh, Hatem Mousselly. Search-based automatic image annotation using geotagged community photos. Diss. INSA de Lyon, 2014.

[12].Bahrami, S., and M. Saniee Abadeh. "Automatic image annotation using an evolutionary algorithm (IAGA)." *Telecommunications (IST), 2014 7th International Symposium on*. IEEE, 2014.

[13].Rusu, Mariana, and Marius-Dan Zbancioc. "Automated identification of objects based on normalized cross-correlation and genetic algorithm." *E-Health and Bioengineering Conference (EHB), 2015*. IEEE, 2015.

[14].Nwana, Amandianeze O., and Tsuhan Chen. "Who ordered this?: Exploiting implicit user tag order preferences for personalized image tagging." *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016.

[15].Sandnes, Frode Eika. "Geo-Spatial Tagging of Image Collections using Temporal Camera Usage Dynamics." *Pervasive Systems, Algorithms, and Networks (ISPAN), 2009 10th International Symposium on*. IEEE, 2009.

[16]. Evertsen, Martin Hætta. *Automatic Image Tagging based on Context Information.* MS thesis. Universitetet i Tromsø, 2014.