Image Tagging and Classification

Dishant Mohite, Avinash Monde

Department of Nuclear Information Technology

PVPPCOE,MUMBAI

University of Mumbai

Namrata Musale, Ramesh Holkar

Department of Information Technology

PVPPCOE,MUMBAI

University of Mumbai

Abstract - **The task of effective image tagging typically consists of two stages, which involve initial image tagging and subsequent tag refinement.Image tagging attempts to label an image with one or more human-friendly textual concepts to reflect the visual content of the image .The resultant tags constitute the tag list for this image. Note that image tagging can be done manually by a human, or automatically by an algorithm Image tag refinement aims to remove imprecise tags and supplement incomplete tags, since the tags in a tag list may be imprecise for that image, and some relevant tags may be missing from the tag list.We will create a software that will take images as user inputs and tag them based on the features of the image using Auto-Tagging Feature .First we will create a training set of data and train our software to tag it based on auto-tagging algorithms and then the software will automatically tag the images based on the features included in image when a user uploads or give url of an image.Also we will include the Image Classification which will classify the images uploaded by the user based on the features or what the image is about So here we will use classifier algorithms which are based on Neural Networks And the Clustering algorithms to classify the images based on pixels and shapes and sizes and also we will classify them on basis of texture, geometry and context. This will help the users and organizations to easily classify their images automatically instead which have taken lot of time doing manually.**

Index Terms:- Automatic image tagging,CNN.

I. INTRODUCTION

Over the past decade the number of images being captured and shared has grown enormously. There are several factors behind this remarkable trend. In the modern age it is now commonplace for private individuals to own at least one digital camera, either attached to a mobile phone, or as a separate device in its own right . The ease with which digital cameras allow people to capture, edit, store and share high quality images in comparison to the old film cameras. This factor, coupled with the low cost of memory and hard disk drives, has undoubtedly been a key driver behind the growth of personal image archives. Furthermore, the popularity of social networking websites such as Facebook and Myspace, alongside image sharing websites such as Flickr. So for this reason image tagging is very important because as there are enumerous images so with helps of tags based on their factors they will be easily classified.[4] Prediction can be used to automatically recognize the value of a parameter.Prediction is used in situations where there are multiple possibilities to a single problem.There are many types of prediction algorithms like Bayesian.

Classification. When the data are being used to predict a category, supervised learning is also called classification. This is the case when assigning an image as a picture of either a 'cat' or a 'dog'. When there are only two choices, it's called two-class or binomial classification.
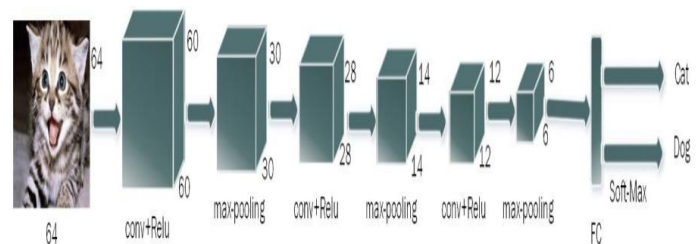
System Diagram

II. LITERATURE SURVEY

Introduction

Image annotation, or prediction of multiple tags for an image,is a challenging task. Most current algorithms are based on large sets of handcrafted features. Deep convolutional neural networks have recently outperformed humans in image classification, and these networks can be used to extract features highly predictive of an image's tags. In this study, we analyze semantic information in features derived from two pre-trained deep network classifiers by evaluating their performance in nearest neighbor-based approaches to tag prediction. We generally exceed performance of the manual features when using the deep features. We also find complementary information in the manual and deep features when used in combination for image annotation.[2]

Existing System

SIFT is used to Automatically assigning mutiple tags to the image. Dataset used isIAPR-TC12 which has They have used Pretrained deep cnn classifiers :AlexNet& VGG-16 [2] which Uses Nearest Neighbor image annotation.The IAPR-TC12 benchmark dataset consists of natural scene imagery and a vocabulary of 291 words The same training and test split was used as in the preceding literature ,with the training and test sets consisting of 17665 and 1962 images, respecttively. 4096 features for fully connected layer used which performs better than other layers.[2]

Algorithms used[2]:

1. TagProp:- This algorithm learns rank of nearest neighbor images with probability model for occurrence of tag

2. 2KPNN: In the 2PKNN algorithm, images are first grouped according to their tags. An image's neighbors consist of the K-closest images from each semantic or tag group, accounting for label imbalance.

Computed precision Metric: Find proportion of images to predict to have tag t that truly have tag in test set

Comparison of Deep and Manual features

1. Regenrated 15 manual features and found results to be highly consistent with previous ones
2. Tag prop produced better accuracy than 2KPNN with AlexNet architecture
3. Deep features Provided better results then Manual features

III. PROPOSED SYSTEM

In our proposed system we have used Cloud services to reduce the requirement of high processing power on the normal computer infrastructure. Also used multiple CNN layers instead of limited number of layers used so that the accuracy will be increased for classification and tagging.We have incorporated some classification architecture which are pre-classified or pre-trained for that purpose so that accuracy will be increased.We have used the SIFT for Image Annotation for directly annotation of tags and tag prop algorithm for same purpose of automatically tagging image based on their characteristics.

Jupyter Notebook:
Jupyter Notebook is an open-source web application which allows us to create documents.Jupyter helps to provide a environment that allows us to see the code,execute the code without interrupting and leaving the environment. This makes it a handy tool for performing end to end data science workflows:- building and training machine learning models, visualizing data, cleaning, and other uses.Because we can execute the code independently in each cells independently.This allows the user to test a specific block of code in a project without having to execute the code from the start of the script.[5]

Libraries
Keras:
Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow.It was developed to support the execution of code at faster rate *Being able to go from idea to result with the least possible delay is key to doing good research.* Designed to enable fast experimentation with neural networks, it focuses on being user-friendly, modular, and extensible. It offers a higher-level, more inovative set of abstractions that make it easy to develop deep learning models regardless of the computational backend used[7].

TagProp Algorithm:
The TagProp algorithm learns rank- or distance-based weights of nearest neighbor images as part of a joint probability model for the occurrence of tags associated with images. The weight of training image j for query image i is

$\pi i,j = e{-d\theta(i,j)}$ j $e{-d\theta(i,j)}$ where, in the multi-distance case, θ is a vector of coefficients,
d θ(i, j) = θ
di,j , and di,j is the vector of distances between images i and j. In the σ variant of the algorithm, prediction of rare tags (a known impediment to prediction recall) can be enhanced with a logistic discriminant model.[2]

AlexNet Architecture:
AlexNet is the name of a convolutional neural network, designed by Alex KrizhevskyIt contains 5 convolutional layers and 3 fully connected layers. Relu is applied after very convolutional and fully connected layer. Dropout is applied before the first and the second fully connected year. The image size in the following architecutre chart should be 227 * 227 instead of 224 * 224, as it is pointed out in his famous CS231n Course. More insterestingly, the input size is 224 * 224 with 2 padding in the pytorch torch vision. The output width and height should be (224–11+4)/4 + 1=55.25!. It consisted 11x11, 5x5,3x3, convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. It attached ReLU activations after every convolutional and fully-connected layer. AlexNet was trained for 6 days simultaneously on two Geforce GTX 1060ti GPUs which is the reason for why their network is split into two pipelines[9].

IV. CONCLUSIONS

This work presents a classification process using the deep learning architecture which is frequently used to solve image processing problems. Although existing classification processes are considered successful, their use in fields such as safety and health is critical, where it is critical to find the right one. Therefore, methods are needed to improve the accuracy rate. Deep learning proved itself by solving many machine learning problems. The use of deep learning architectures is important in the methods developed at this point. In our future work, we will improve the proposed model in terms of speed performance and accuracy. We will perform facial recognition by pointing out the need to increase safety recently with the model we have developed. We have clearly demonstrated that features derived from a deep convolutional neural network match or exceed image annotation performance using larger manual feature sets. We have also provided evidence of complementary information in both the deep and manual features, suggesting they could be used in conjunction to enhance predictive performance, depending on the dataset under study. We note that we used the pre-trained networks as feature transforms without back-propagating tag prediction errors through the network. As part of current and future work, we are developing deep learning frameworks that fully integrate multimedia feature extraction with annotation. Taken together, this analysis supportsmore widespread adoption and further investigation of deeply learned feature representations in multimedia labeling tasks.

V.REFERNCES

[1] EmineCengil,AhmetCinar,ZaferGuler,” A GPU based convolutional neural network approach for image classification ”department of computer engineering,firatuniversity,turkey,©2017 IEEE

[2] Michael B. Mayhew∗, Barry chen ,”Assessing Semantic Information In Convolutional Neural Network Representations Of Images Via Image Annotation”Lawrence Livermore National LaboratoryComputational Engineering Division Livermore, CA USA, IEEE©2016

[3] Jianlongfu and yongrui.,” Advances in deep learning approaches for image tagging” ©2017 IEEE

[4] Sean Moran,” Automatic Image Tagging” ©2009 IEEE

[5] https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html

[6] https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-   totally-from-a-beginners-perspective-be8dbaf22dd8

[7] https://keras.io/

[8] https://www.geeksforgeeks.org/introduction-convolution-neural-network/

[9] https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637