ijRASET

International Journal For Research in
Applied Science and Engineering Technology

Transfer Learning Based Multi-Label Classification of Images

K.S.Anand[1], M.Poonkodi [2], C. G. Rohith Kumar Gupta[3], Raj Kshatriya Kshaktawat[4]

[1, 2, 3, 4]Department of Computer Science & Engineering, SRM Institute of Science & Technology, Tamilnadu , India

Abstract: Multi-label classification task is concerned with classifying an image into one or more classes(categories) based on the content of the image itself. Multi-label classification is different from binary or multi-class classification wherein the aim of the classifier built is to classify the image into a single class from a set number of classes. Existing methods utilize feature extraction techniques such as colour histograms, SIFT which are limited by their representational ability. We propose to overcome this problem by leveraging the rich features that can be extracted from CNN that have been trained on million images. The features are then fed into an Artificial Neural Net, which is trained on the image features and multi-label tags. By utilising transfer learning, we harness the feature representational ability combined with reduced training time. We benchmark the model with dataset obtained from Flickr (FLICKR-25K). The evaluation metrics utilised here include mAP, Training accuracy and Training Loss.
Keywords: Multi-label classification, Histogram, Transfer learning, CNN, SIFT

I. INTRODUCTION

Classification in machine-learning is a task in which the aim is to classify an instance into either a single class or multiple classes from a set of disjoint classes. The former is an example of a multi-class classification and the latter is an example of multi-label classification. Binary classification and /multi-label classification can be thought of as a special case of multi-label classification, where in the number of classes associated with the instance is one. Due to recent realization that there is omnipresence of multi-label classification tasks in the real world there is an increase in the number of researches focused on finding better ways to perform multi-label classification with greater accuracy. Fig 1 depicts an example of Multi-label classification.

Fig 1. An Example of Multi-label classification

With contrast to single-label classification, where in an instance is associated to single class from a set of classes L, multi-label classification aims to associate an instance to a set of classes X which in turn is a subset of classes L. Due to effort and the generality needed in classifying instances to multi-labels the training needed to do so is difficult. Multi-label classification has many applications in the field of text categorization, image tagging etc. In general, the content of image is generally identified by using features extracted from them. Earlier, the features used were colour histograms, histogram of gradients (HoG), texture of the images, Scale-Invariant-Feature-Transformation (SIFT) [1] etc., they lack the representational ability.

Due to recent advances in the field of deep learning, and the increase in the computational ability, high level features can be extracted easily from pre-trained neural nets. The neural net that can be utilized for this purpose is generally trained on millions of images. By leveraging this pre-trained neural net, we can extract high level features from these nets that can be leveraged for better training accuracy. This method of extracting features from a neural net is called as transfer learning [2] . This architecture is depicted in Fig 2. By utilizing transfer learning training time of a neural net is reduced, without compromising the accuracy of the model. Deep learning paradigm is the ability of the neural network to learn from the images directly, without the usage of handcrafted image features. In this paper we look on the work carried out previously and propose a new method of image tagging based on transfer learning.
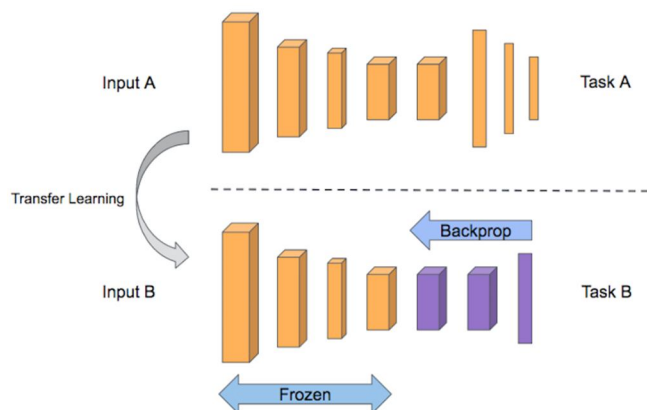
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Fig 2 Architecture of Transfer learning

II. RELATED WORKS

A. Multi-label classification

In general, Classification in machine learning is defined by Andre et al [3] as "Given a set $\{X_i, Y_i\}$ the aim of the classifier is to find a mapping function that associates each instance X_i to class Y_i, where i =1,2, 3, …., n". Multi-label classification in machine learning is a classification task in which the aim is to classify an instance into more than one class i.e., to associate each instance xi to a set of labels Y (where Y > 2). This is different when compared to binary classification, where the aim is to classify an instance into any one of the 2 classes. Fig 3 depicts an example of multi-labelled data.

Instances	Label1	Label2	Label3	Label4
u_1	l_1			l_4
u_2			l_3	l_4
u_3	l_1		l_3	
u_4		l_2	l_3	
u_5		l_2	l_3	l_4

Fig 3 An example Multi-labelled data

B. Transformation And Evaluation Metrics For Multi-Label Classification

Multi-label classification problems are generally transformed before they are solved. The transformation methods available for multi-label classification are transforming into binary classification, transforming into multi-class classification, transforming into ensemble methods. In, transformation of multi-label classification into binary classification problem each class in the dataset has a binary classifier, which classifies whether that particular class is present or not. Examples of this kind of classifier are One vs All, and One vs Rest. In transformation of multi-label classification into multi-class classifier, multiple multi-class classifiers are trained based on the label powerset of the dataset i.e., each powerset of the label incorporates a multi-class classifier. In ensemble methods [4], multiple classifiers are trained and the output of the ensemble is decided based on the voting made by different classifiers. If a class has a pre-requisite number of votes, then it is classified as being present in the instance.

Multi-label classifiers are evaluated differently when compared to other counter-parts because the instance is classified into having multiple classes. In normal classification problems the result given is not partially correct, whereas it is the case in multi-label classification. Some of the metrics used are: Jaccard index, Precision, Recall and F1 score. These metrics are calculated by equations (1-4) , where T is the actual label set and To is the label set predicted by the classifier.

$$Jaccard\ Index: \frac{|T \cap To|}{|T \cup To|} \qquad (1)$$

$$Precision: \frac{|T \cap To|}{|To|} \qquad (2)$$

$$Recall: \frac{|T \cap To|}{|T|} \qquad (3)$$

$$F1\ Score: \frac{2\,|T \cap To|}{|T|+|To|} \qquad (4)$$

III. DATASET DESCRIPTION

The dataset used to build the classifier is the MIRFLICKR-25000 dataset [5] which consists of 25000 images with their corresponding tags and ground truth tags. There are 24 ground truth classes and 1386 tags. The distribution of images of the top 15 classes is described in the following figure (Fig 4).

Fig 4. Distribution of images for top 15 classes

The ground truth classes in the FLICKR-25K dataset is tabulated in table below. (Table 1)

TABLE 1 TABULATION OF VARIOUS CLASSES IN DATASET

CLASSES					
Animals	Baby	Bird	Car	Clouds	Dog
Female	Flower+	Food	Indoor	Lake	Male
Night	People	Plant life	Portrait	River	Sea
Sky	Structures	Sunset	Transport	Tree	water

The classes have semantic overlapping i.e., an image of tree will be tagged as [tree, plant life]. Therefore, it is, important to understand this semantic overlapping while building the multi-label classifier [6]. The semantic overlapping in the dataset is mapped as a class being a general topic of a subclass or vice-versa. The classes that have a semantic overlapping are as follows:

TABLE 2 RELATIONSHIP BETWEEN GENERAL TOPIC & SUBTOPICS IN DATASET

GENERAL TOPIC	SUB TOPIC
Sky	Clouds
Water	Sea/ocean, river, lake
People	Portrait, boy/girl, man/woman, baby
Plant life	Tree, flower
Animals	Dog, bird
Structures	Architecture, building, house, city, bridge, road
Transport	Car

IV. PROPOSED WORK

The proposed model for multi-label classification involves using a pre-trained neural net to extract the features from the images, which in turn is fed into an artificial neural net which is trained in these features to find the mapping between the image features and the multi-label classes. During testing phase, the test image is passed through the same pre-trained neural net which extracts the features, which is fed into the trained neural net which then predicts the classes for the image. The artificial neural net uses Adam optimizer and accuracy, top-k-accuracy as a validating measure. The neural net utilizes an Adam optimizer [7] with a learning rate of 0.001.

$$W_t = W_{t-1} - \eta \frac{m_t}{\sqrt{v_t + \epsilon}} \qquad (5)$$

The changing of weights by the Adam optimizer is carried out based on the formula (5) given above, where W_t is the weight of the current step, W_{t-1} is the weight of previous step, η is the step size, m_t and v_t are bias corrected first and second moment values. The architecture of the proposed model is picturised in the following figure (Fig 5).
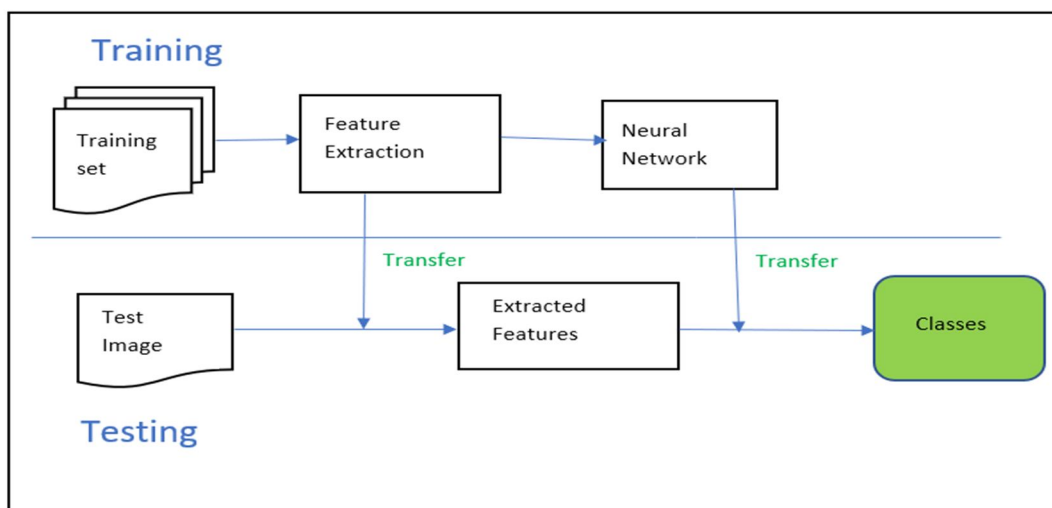

Fig 5. Architecture of the proposed system

V. RESULT

Loss and accuracy are the two measures that we monitor through out the training of the neural net. The loss and accuracy of neural net during training is captured as a loss (Fig 6) and accuracy graph (Fig 6). We compare our neural network model's mean average precision score with pre-existing systems, which shows that our model outperforms them (Table 3).
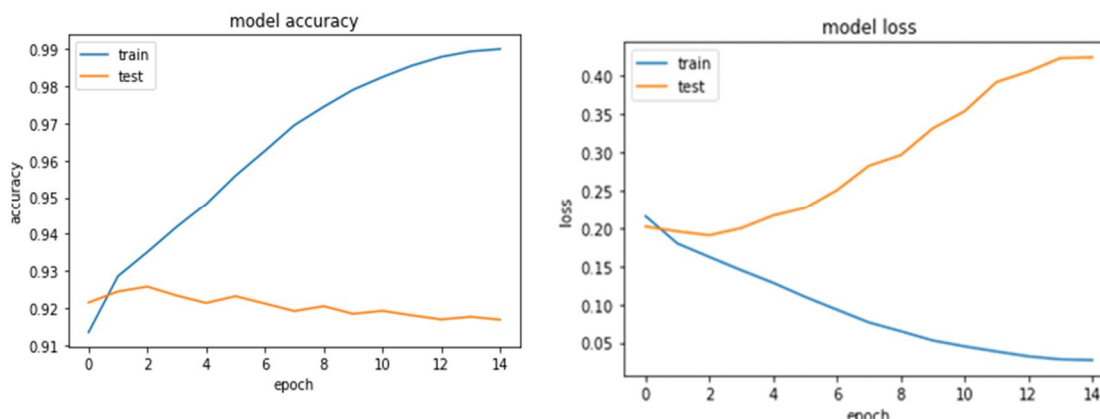

Fig 6. Graaph depicting variation of model accuracy and model loss during training

TABLE 3 COMPARISON OF RESULTS WITH EXISTING SYSTEMS

Method	mAP Score
BoW + Tagprop [8]	0.33
CNN + KNN [9]	0.63
CNN + Tagprop [10]	0.65
ResNet-50 + NN	*0.72*

VI. CONCLUSION

In this paper, we propose a transfer-learning based approach for multi-label classification of images, utilising one-hot encoding. This approach outperforms other existing systems and reduces the training time of the neural net drastically. The mean average precision score obtained is significantly higher than those of the existing systems. By utilising different pre-trained neural networks one can improve the mAP score obtained. The multi-label classification of images can be improved by using various techniques such as Word vectorisation [11] , Bag of words (BoW) [12] , Continuous bag of words (C-BoW) [13-14] etc.

REFERENCES

[1] Cheung, Warren, and Ghassan Hamarneh. "$ n $-SIFT: $ n $-Dimensional Scale Invariant Feature Transform." IEEE Transactions on Image Processing 18.9 (2009): 2012-2021.

[2] Torrey, Lisa, and Jude Shavlik. "Transfer learning." Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010. 242-264.

[3] de Carvalho, André CPLF, and Alex A. Freitas. "A tutorial on multi-label classification techniques." Foundations of Computational Intelligence Volume 5. Springer, Berlin, Heidelberg, 2009. 177-195.

[4] Bhatia, Kush, et al. "Sparse local embeddings for extreme multi-label classification." Advances in neural information processing systems. 2015.

[5] Huiskes, Mark J., and Michael S. Lew. "The MIR flickr retrieval evaluation." Proceedings of the 1st ACM international conference on Multimedia information retrieval. ACM, 2008.

[6] Wang, Jiang, et al. "Cnn-rnn: A unified framework for multi-label image classification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[7] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[8] Li, Xirong, Cees GM Snoek, and Marcel Worring. "Learning tag relevance by neighbor voting for social image retrieval." Proceedings of the 1st ACM international conference on Multimedia information retrieval. ACM, 2008.

[9] Makadia, Ameesh, Vladimir Pavlovic, and Sanjiv Kumar. "Baselines for image annotation." International Journal of Computer Vision 90.1 (2010): 88-105.

[10] Branson, Steve, et al. "Bird species categorization using pose normalized deep convolutional nets." arXiv preprint arXiv:1406.2952 (2014).

[11] Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[12] Zhang, Yin, Rong Jin, and Zhi-Hua Zhou. "Understanding bag-of-words model: a statistical framework." International Journal of Machine Learning and Cybernetics 1.1-4 (2010): 43-52.

[13] Bansal, Mohit, Kevin Gimpel, and Karen Livescu. "Tailoring continuous word representations for dependency parsing." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vol. 2. 2014.

[14] Sridhar, Vivek Kumar Rangarajan. "Unsupervised topic modeling for short texts using distributed representations of words." Proceedings of the 1st workshop on vector space modeling for natural language processing. 2015.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)