JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Literature Review of Multiclass Image Classification Based on Inception-v3 Transfer Learning Model

Pooja Kushwah, Monika Bhatnagar

ABSTRACT Transfer learning is the reuse of a pre-trained model for a new problem, it is very popular nowadays in deep learning because it can train deep neural networks with relatively little data, and it is very useful in data science because of most real problems., you don't have millions of data points marked to train these complex models. Let's take a look at what transfer learning is, how it works, why and when to use it. Includes several resources for models that have been previously trained in learning transfers for example, when you train the classifier to predict whether an image contains food, you can use the knowledge gained during training to recognize drinks, for example, if you trained a simple classifier to predict, if the image includes a backpack, you can use the knowledge gained by the model during training.

INDEX TERMS: Convolutional neural network, data augmentation, deep learning, multiple food recognition, transfer learning.

I.INTRODUCTION

Several approaches have been done to classify food from images. In previous years many feature based models were used to classify food images. SCD , EFD , GFD and LBP are the common features that has been used to classify food images. In modern literature there are neural networks especially convolutional neural networks have been used to classify food images.

1. Feature Representation Transfer

Feature representation level knowledge transfer is a popular transfer learning category that maps the target domain to the source domains exploiting a set of meticulously manufactured features. Through this type of feature representation level knowledge transfer, data divergence between the target domain and the source domains can be significantly reduced so that the performance of the task in the target domain is improved. Most existing features are designed for specific domains and would not perform optimally across different data types. Thus, we review the feature level knowledge transfer techniques according to two data types: 1) cross-domain knowledge transfer and 2) cross-view knowledge transfer.

Feature recognition is quite different from various object recognition algorithms. These algorithms are based on one type of feature-edge. Since the world is full of edges that look pretty much the same, the set of edges extracted from the image must first find the mapping (matching) from the edge of the image to the edge of the model before making a direct comparison with the set of edges extracted from the model object. In each case, meaningful calculations are required.

perception to be viewed as a distribution problem. It proceeds through the computing and various properties of the input image and combines them into feature vectors. An object model is a set of feature vectors related to a set of representative images of an object. New images are categorized by calculating the image's feature vector and comparing it directly with the model vector. If the object model contains a feature vector that is closest to the feature vector of the image, then the image is cited as an object instance.

Scale Invariant Feature Transformation (SIFT) is a computer vision algorithm for finding and describing local features in an image. Glossary tree 42 implemented with the closest adjacent food category k and 1453 images. For distance measurements, the Euclidean distance (L2 norm) of the L1 norm DCD function is selected for the SCR, EFD, and GFD characteristics. The combination of DCD, MDSIFT, SCD and SIFT.

functions resulted in 64.5% Top 1 accuracy and 84.2% Top 4 accuracy. In the SVM classifier, a method has been proposed to use the SIFT and LBP functions with a PFI data set, the SIFT function is used to find and describe local features in an image, and LBP is a kind of visual descriptor, many LBPs are easy to calculate and are sensitive to lighting changes Unaffected Support Vector Machine is a supervised learning model with associated learning algorithms for analyzing data used for classification and regression analysis. There is a way to classify food images using spherical surfaces. Machines that support vector machines efficiently perform nonlinear classification using kernel tricks, apply this method to a food log data set consisting of 6512 images, and split using an FCM algorithm similar to the k-means clustering algorithm. Can be applied to food images. The coefficient is assigned. For each data point in the cluster, the centroid is calculated randomly for each cluster and a coefficient is calculated for each data point. After applying the FCM to segment the food image, I used a spherical support and the accuracy for classifying the Food 101 data set is 95 Classifier. The Random Forest or the Random Forests is a collaborative way of dividing, retreating, and other activities that involve building a series of decision making trees during training and taking classes in class mode (phases) or intermediate predictions (retreat). with an accuracy of 50.76 using the RFDC method.

Learning is a tool for improving the performance of model domain targets in that case the target domain label is not enough, otherwise the moving knowledge is meaningless. So far, most studies of learning are focused only on a small scale of data, which cannot also reflect the potential of learning on the machine regularly learning techniques. Future challenges of learning should be in two aspects:

1) how to exploit information that will be useful for regional targets from high noise source data domains and 2) How to expand the current transfer of learning methods to deal with large scale of Data Domain sources.

2. Deep Learning Based Model

Nowadays in-depth reading is very popular in computer vision. Deep reading (also known as deep reading or sequential reading) is part of a wider family of machine learning methods based on the

representation of learning data, unlike task algorithms. such as deep neural networks, deep belief networks and duplicate neural networks used in fields including computer vision, speech recognition, natural language processing, sound recognition, social network filtering, machine translation, bioinformatics, drug production and board game programs, where they produce comparable results and, in some cases, superior to human professionals. In recent publications there are many methods that have used the deep convolution network neural network to edit food images. A neural network of food image classification. In-depth study was used to classify the UEC-256 food image analysis of a computer-assisted testing program. CNN was used to classify food images in order from the food-11 dataset to build a dietary management system.

A pre-trained deep neural network was applied. Deep convolutional neural networks were pretrained on ImageNet with 1000 food-related categories then fine tuned. To classify the UEC-FOOD100 dataset, we achieved 78.77% top-1 accuracy.

We used Google Net to classify Thai fast-food images on the TFF food dataset. We achieved 88.33° the food-11 dataset. We find that the a deep learning approach with an accuracy of 12% on the proposed approach, 80.51% on the Caffe net, and 82.07 on the Alex net yields better results than traditional feature-based models with larger datasets. 11 datasets. I tried a convolutional neural network built from scratch and transfer learning using the Inception V3 model.

3. Inception- Overview

In this paper, Inception, it was developed according to the Google Net architecture seen in ILSVRC 2014. It is also inspired by the method based on the primate visual cortex dictated by Serre et al. , which can capture scales. many sizes, one of the key criteria of fund forming architecture, is the adaptation of the network "in the network" method. Lin et al, which increases the power of artificial neural networks. the reduction in size is 1 × 1. The purpose of fund architecture is to reduce the use of resources to classify. Accurate images use deep learning. They focus on finding the best position between traditional methods of optimization. This increases the size and depth, and uses sparsity in layers depending on the theoretical area set by Arora et al.. It itself can pay a lot of calculated resources for deep learning systems such as establishing funds which use filters, in their 22 layers architecture, which is the main goal to achieve them, emphasizing the approach of Arora et al.to generate a correlation statistical analysis to generate groups of higher correlation to feed forward to the next layer. And they took the idea of multiscale analysis of visual information in their 1 × 1 , 3 × 3 and 5 × 5 convolution layers. All of these layers then go through dimension reduction to end up in 1 × 1 convolutions .

The Inception architecture used in ILSVRC 2014 had the following structure as denoted by Szegedy et al.: • An average pooling layer having 5 × 5 filter size and stride 3.

• A 1 × 1 layer with 128 filters for dimensionreduction and rectified linear activation.

• A fully connected layer having1024 units andrectified linear activation.

LITERATURE REVIEW

To categorize food from photos, several approaches have been used. Many feature-based methods were tried to classify food photographs in earlier years. The common features that have been used to identify food photos include SCD, EFD, GFD, and LBP. Neural networks, particularly convolutional neural networks, have been used to categorize food photos in recent literature.

Based on the study of numerous authors' work, the following information regarding the proposed '**Food Classification Model'**:

1. Burkapalli, V.C. and Patil, P.C[1] reviewed the Google Inception-V3 model 1; as a base, a fully linked layer is developed on top of it to optimise the categorization process. Convolution layers are capable of learning enough on their own convolution kernel to yield tensor outputs during the model development process.

In addition, before the classification step, the segmented features are concatenated with our own model. It improves the capability of key traits and is used in the food categorization process.

2. Z. Zong, D. T. Nguyen, P. Ogunbona, and W. Li[10] undertook a systematic examination of key scientific databases, including interdisciplinary databases (such as Scopus) and academic databases in the field of computer science that focus on image understanding topics (i.e., recognition, analysis, retrieval).

3. Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, [7] investigated food picture analysis variables and their combinations, as well as a classification method based on k-nearest neighbors and vocabulary trees. The system is tested using a food image dataset that includes 1453 photographs of eating occasions in 42 food categories, taken by 45 people in natural eating situations. The classification system described in the previously reported work [1] is tested using the same image collection. The use of our mix of features and vocabulary trees for classification improves food classification performance by roughly 22% for Top 1 classification accuracy and 10% for Top 4 classification accuracy, according to experimental results.

4. K. Yanai and Y. Kawano,[15] emphasized on the food image recognition , five-layer CNN and data expansion techniques were used to enhance the size of training images, resulting in a 90% increase in accuracy.

LIMITATIONS the model is to be developed and implemented according to the framework and datasets as stated. We shall use Convolution Neural Network as the algorithm to conduct the training and

implementation of the deep learning model to make predictions for food classification. After data pre-processing, the chosen datasets, Food-101 . We use a subset of the four food categories [Chicken Curry, Hamburger, Omelet, Waffle]. shall be used to train the model where patterns and characteristics of the food images are distinguished over multiple passes of the neural network. Once that is completed, testing is done to ensure the model achieves accuracy in prediction and fit for deployment in a food classifier web application.In the future, the ability to evaluate, analyse, and use picture recognition technologies was required. Visual data comprised the majority of his digital marketing. Otherwise, the company would have overlooked a wealth of information. It's possible that media monitoring will go unnoticed. Artificial intelligence and image recognition make it easier for marketers to identify photos on social media without explicit textual mention, so you don't lose out on a great opportunity to learn and engage with your clients.

The need to have a system that measures daily food intake for healthy diet is crucial due to the insufficient knowledge of diet and calorie requirements. In addition, correct food recognition is considered challenge. Hence, we proposed a measurement.

PROPOSED METHODOLOGY

In our thesis we used a part of food 101 dataset for our research. We developed a CNN from scratch to classify food images.

We also used transfer learning from Inception v3 model which was pre-trained with ImageNet method in our work. This chapter describes the model of our

convolutional neural networks and transfer learning method in details.

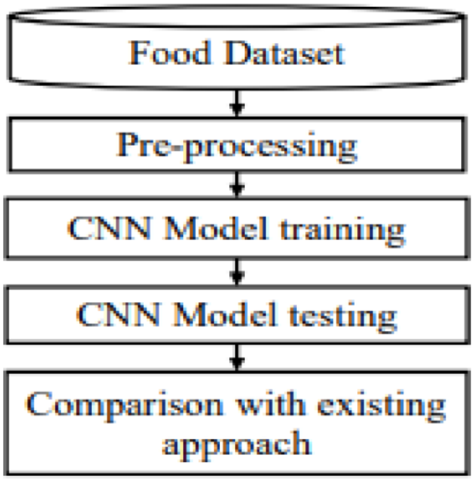

Fig.: Proposed Model Approach

We employed four layers in the suggested system: convolution layer, relu layer, pooling

layer, and fully linked layer.Each convolutional layer is followed by a pooling layer, which reduces the image dimension while maintaining spatial invariance. Hence decreases the CNN network's computation cost The max-pooling filter size in our architecture is 3. A total of 100 epochs have been used to train the CNN model, with the most prevalent features being chosen;by choosing the highest feature value from the previous layers After collecting and

converting all of the features. Fully connected (FC) layer is used to map features and classify images into correct categories.

CONCLUSION

Other neural network models such as recurrent neural network , dilated convolutional neural network , etc. can be applied to classify food images. Convolution neural network models take time for computation but once the model is trained it can be easily used for classification.

This study aims to categorize Indian food photographs into their appropriate categories. The suggested software model is based on machine learning, and it detects the food image that the user uploads as an input, processes it, recognizes it, and estimates calories from the anticipated image. People are more willing than ever to capture, upload, and share food photographs on social media sites like Instagram and Facebook. As a result, it is easier to find more food-related data (pictures and videos). As a result, consumers may better monitor their diets, and the consumption of manual paper is reduced. We can also improve our research by

applying feature-based models which will take less computational time. We will try to make an android application in future to detect food from images. The CNN model should be trained on more category of foods to work efficiently with social media to classify foods because there are thousands more category of foods worldwide. If the CNN can be trained with more category of foods then it will help social media platform to classify foods efficiently even if the food is not labeled in caption. It will also help the restaurants to advertise their targeted audience efficiently. It will also help consumer to choose restaurants with similar kind of food We will also try to build a recommender system that will suggest restaurants to the people according to their food preferences.

In this thesis we analyze that our system model performance is significantly high, and according to our simulation, we have analyzed that it requires a high-performance system for large number of data sets. data is non-linear, but it requires more - more computation time to train the model; However, the model's performance matters a lot, so once the system model is properly trained, it will take very little time to produce systematic results. All images under review are appropriately pre-processed, and all image types are tested using the CNN model. According to the analysis done, we can conclude that our proposed CNN model

matches the food image.

II.REFERENCES

[1] Burkapalli, V.C. and Patil, P.C., TRANSFER LEARNING: INCEPTION-V3 BASED CUSTOM CLASSIFICATION APPROACH FOR FOOD IMAGES.

[2] C. Szegedy et al., "Going deeper with convolutions," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston,MApp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[3] RCBolles and R.A. Cain, "Recognizing and locating partially visible objects: The local-feature-focus method," vol.1,no. 3, , pp.57–82.[Online]. Available: https://doi.org/10.1177/027836498200100304

[4] D. Scherer, A. Mu¨ller, and S. Behnke, "Evaluation of pooling operations in convo- lutional architectures for object recognition," pp. 92–101, 2010.

[5] N. Srivastava,G.Hinton, A. Krizhevsky, I. Sutskever, andR. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15,pp[Online].Available: http://jmlr.org/papers/v15/srivastava14a.html

[6] C. Bishop, Pattern Recognition and Machine Learning.Springer-Verlag, 2006.

[7] Y. He, C. Xu, N. Khanna, C. J.Boushey, and E. J. Delp, "Analysis of food im- ages: Features and classification," in2014 IEEE International Conference on Image Processing (ICIP),, pp. 2744–2748.

[8] D. chen He and L. Wang, "Texture unit, texture spectrum, and texture analysis,"IEEE Transactions on Geoscience and Remote Sensing, vol. 28, no. 4, pp. 509–512, July 1990.

[9] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food im- ages: Features and classification," in IEEE International Conference on ImageProcessing (ICIP), , pp. 2744–2748.

[10] Z. Zong,D.T.Nguyen, P.Ogunbona, and W. Li, "On the combination of local texture and global structure for food classification," IEEE International Symposium on Multimedia,