



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: II Month of publication: February 2021

DOI: <https://doi.org/10.22214/ijraset.2021.33018>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

A Survey on Multiclass Image Classification based on Inception-v3 Transfer Learning Model

Prof. Kanchan V. Warkar¹, Anamika B. Pandey²

^{1,2}M.Tech CSE Department, Bapurao Deshmukh College of Engineering Sewagram

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: Food image, Transfer learning, Inception-v3.

I. INTRODUCTION

Several approaches have been done to classify food from images. In previous years many feature based model is being 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.

A. 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, and instead of trying to get rid of the matching problem, FBR builds a representation of the model that can be compared directly to the image, allowing 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 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 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 domain and 2) How to expand the current transfer of learning methods to deal with large scale of Data Domain sources.

B. 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 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 pretrained deep neural networks were applied. Deep convolutional neural networks were pretrained on ImageNet with 1000 food-related categories than 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° Curacy for 11 classes. Implemented and compared several convolutional neurons. I got 70 network models with the food-11 dataset. We find that the 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.

C. 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 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. 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 use 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.:

- 1) An average pooling layer having 5×5 filter size and stride 3.
- 2) A 1×1 layer with 128 filters for dimension reduction and rectified linear activation.
- 3) A fully connected layer having 1024 units and rectified linear activation.
- 4) A dropout layer having 70% ratio of dropped outputs.

II. DESCRIPTION OF PLANNED SYSTEM

In the proposed system we used a part of the original food-101 dataset with 101 food categories. All images are rescaled to a maximum side length of 512 pixels. Use a subset of the four food categories [Chicken Curry, Hamburger, Omelet, Waffle]. The data consists of three main subfolders: training, validation, and testing. The training data consists of 1000 images per class, with up to 500 validation images and up to 500 test images per class. The dataset has not been (intentionally) cleaned up and therefore contains some noise. It is mainly displayed in dark colors and in some cases has the wrong label. The dataset is not complete, which makes the problem even more difficult. However, it uses the assigned label. 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.

We have taken some image pre-processing technique to increase efficiency to our system. First, we re-sized all our images to 224 x 224 x 3 to increase processing time and also to fit in our convolutional neural network model. After that we applied following image pre-processing techniques.

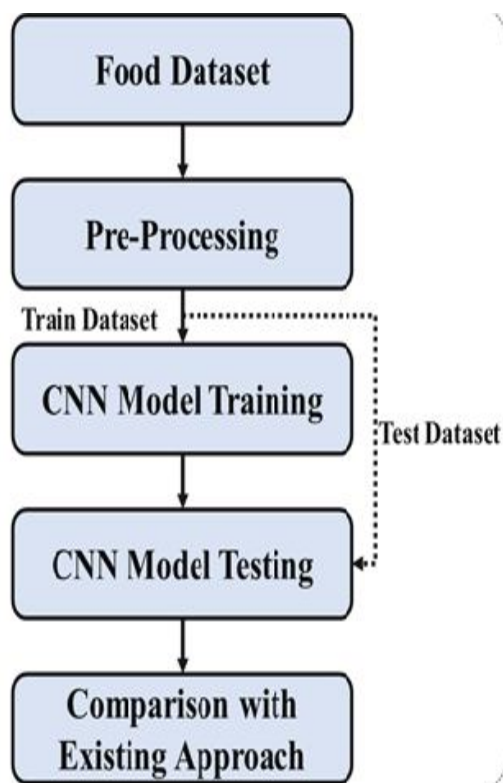


Fig 1: Methodology of food image classification.

Neural Networks often try to detect edges in earlier layers, shapes in the middle layer, and some features specific to tasks in the following layers. It helps leverage the labeled data of the task for which it was originally trained. The model has learned to recognize objects, so we will only retrain the following layers.

During transfer learning, we try to transfer as much knowledge as possible from the previous task that the trained model has to the new task at hand. This knowledge can take many different forms depending on the problem and the data. For example, it could be the way models are constructed, allowing us to more easily define new objects. , a lot of data is needed to train a neural network from scratch but not always have access to that data available - this is where transfer learning becomes useful. because the model has been trained in advance. This is especially valuable in natural language processing since most of the expertise is required to create large-labeled datasets. Also, training time is reduced because it can sometimes take days or even weeks to train a deep neural network from scratch for a complex task.

III. SCOPE AND LIMITATION

This model can be used to classify any food image in just seconds, we can also classify any food image by using this approach. Future work would involve more optimization on hyperparameters and model aspects such as which layers to freeze versus make trainable during transfer learning. Due to computing resource and time constraints, most model implementation decisions were made by examining the convergence of the model and relative metrics from training versus validation but an exhaustive hyperparameter search would have been a more empirical approach.

REFERENCES

- [1] L. Shao, F. Zhu and X. Li, "Transfer Learning for Visual Categorization: A Survey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 5, pp. 1019-1034, May 2015, doi: 10.1109/TNNLS.2014.2330900.
- [2] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [3] Neyshabur, B., Sedghi, H., & Zhang, C. (2020). What is being transferred in transfer learning? ArXiv, abs/2008.11687.
- [4] Tan C., Sun F., Kong T., Zhang W., Yang C., Liu C. (2018) 'A Survey on Deep Transfer Learning'. In: Kůrková V., Manolopoulos Y., Hammer B., Iliadis L., Maglogiannis I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science, vol 11141. Springer, Cham. https://doi.org/10.1007/978-3-030-01424-7_27
- [5] Cheng Wang, Lin Hao ; Xuebo Liu ; Yu Zeng ; Jianwei Chen ; Guokai Zhang et. al. C. Wang et al., "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," in IEEE Access, vol. 7, pp.146533-146541, 2019, doi: 10.1109/ACCESS.2019.2946000



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)