iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.40033

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Diverse Approach on Image Categorization Using Transfer Learning Methods

B. Dhanapriya[1], S. Kavitha[2], K. R. Baskaran[3]
[1]*Department of Information Technology, Kumaraguru College of Technology*
[2]*Department of Information Technology, Kumaraguru College of Technology*
[3]*Department of Computer Science, Kumaraguru College of Technology*

Abstract: ***It is an ingrained ability of humans to recognize and classify an image within a millisecond. This is because since our childhood, the human brain is accustomed to seeing a variety of images from the same category. However image classification in computers is a challenging process. To train computers to recognize and categorize images to a specific category, thousands of images of the same category must be sent, by which the computers can figure out and store the pattern from all the images of that specific category. When an image of the same category is sent again, it will easily recognize the image belongs to a specific class based on the patterns that are stored for that class. The objective of this paper is to explore the different transfer learning techniques that can be used for image classification task with high accuracy.***
Keywords: VGG19, ResNet, Densenet, Inceptionv3, Xception

I. INTRODUCTION

The brain assists humans in visualizing whereas computer vision is the field that enables computers to visualize images and videos. Computer vision always seeks to mimic the brain in order to allow computers to understand images and derive meaning from them, just as we do with our brains.Computer vision has reached to such a peak in today's world is because it plays a vital part in self-driving cars and also in augmented reality field.The field of computer vision is one that has become so important that it is used in almost all industries like health care, security etc. A computer reads an image by dividing the image into several pixels. In the image, a pixel is the tiniest dot that represents a particular colour. Computers can comprehend only 0's and 1's.So,to convert a pixel into a number it uses various codes, the most common of which is RGB colour code. Since RGB colour code is used in pixels, each pixel can be thought of as a combination of these three colours, red, green and blue .RGB colour codes have values ranging from 0 to 255, so each pixel has a value ranging from 0 to 255.Thus computers can understand an image by grouping all these pixel values.
In order to extract significant information from images after the computers have been able to read them, the computer vision field involves several tasks. Below are the four different tasks in this field of computer vision which is done in order to do deduce precise information from images.

A. Image Classification

Classification of images comes under supervised learning and it involves placing the image in a category to which it belongs. Image classification involves assigning a single tag to the entire image.

B. Object Localization

Multiple objects are present in images, so the next task is to determine where each of these objects is located.As a result of object localization, all objects present in that image is identified and a rectangular box is drawn around them in order to highlight them. The rectangular box around the objects is known as bounding box.

C. Object Detection

The object detection process combines image classification and object localization tasks by finding all the objects present in the image, drawing a rectangular box around it and assigning a tag to each object based on the class to which it belongs.

D. Segmentation

Segmentation involves dividing an image into smaller segments where each segment is a group of pixels and they are represented by masks. Below are the two different segmentation techniques used in computer vision field.

©IJRASET: All Rights are Reserved

1) *Semantic Segmentation:* Semantic segmentation task involves assigning tag to every pixel of the image to which it belongs to.
2) *Instance Segmentation:* Instance segmentation identifies the objects in the image as unique instances despite belonging to the same class. In this paper we have explored the different transfer learning technique which is used for assigning a single label to the entire image, i.e. to categorize the image to a specific target class.

II. RELATED WORK

According to Monika Bansal [15], a pre-trained model VGG19 has been used to classify images on the Caltech-101 dataset. In this paper they have combined the VGG19 pre trained model with multiple feature extraction methods for classifying the images. With the extracted features, they have attempted to classify the images with several classifiers, but the best result came from combining the extracted features with random forest classifier, and achieved an accuracy of 93.73%.

A new procedure was proposed [14] in which the ResNet pre-trained network is used to detect colorectal cancers. Using ResNet architecture, they classified the image as cancerous or non-cancerous. Authors achieved an accuracy of over 80% using ResNet pre-trained networks.

Authors of the paper [13] used InceptionV3, a special pre-trained network, to identify pulmonary disease. A number of classifiers were used after separating the features, including SVM, softmax, etc., and achieved an accuracy of around 95.41%.

For classifying cancer images, a new pre trained network was proposed in this paper [12] known as Densenet which gives precise results in classification compared to other pre-trained networks like VGG19 and ResNet models.

In this paper [16], an advanced Xception network known as L2MXception is described which helps in the classification of peach diseases in plants. A new network proposed by the author achieved a better accuracy of around 93.85% in detecting peach disease.

III. TRANSFER LEARNING

Pre-trained models are those which are constructed in the notion of providing a solution to certain problems. Pre-trained models are developed by training on massive amount of data so that they have a very high accuracy score.In transfer learning, a pre-trained network is used to provide a solution to a new problem whose nature is similar to the problem the pre-trained network was trained to solve.

The transfer learning process eliminates the need to build a model from scratch, allowing us to make changes to an existing model and use it to solve the specific problem.

Benefits of using transfer learning techniques include improved performance and time efficiency. Transfer learning plays a vital role in several applications like classification of text, reinforcement learning etc.

IV. FACTORS IN CHOOSING PRE-TRAINED MODEL

First, one needs to select appropriate pre-trained model, in order to employ transfer learning techniques to solve the image classification problem. There are number of factors that should be considered when choosing a pre-trained model suitable for image classification.

Below are the few factors that need to be considered in choosing an appropriate pre-trained model for problems.

A. Preferred Output

Selecting a model should be based on the output we seek.

B. Input Dataset

Choosing a model depends on how much data it needs to train on. In the case of a small dataset, we will use a pre-trained model that only has a few layers. In the case of a large dataset we will use a model that has deep layers. So the choice of model for a given problem depends on its dataset.

C. Accuracy Metric

As a result, the model we choose should give high accuracy scores. Therefore, the classification task can be performed on different pre-trained networks and we should pick the model that gives the highest accuracy score.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

V. DIFFERENT PRE-TRAINED MODELS FOR IMAGE CLASSIFICATION TASK

A number of pre-trained networks are used for image classification. The best ones are listed below.

A. VGG-19 Pre-trained model

A pre-trained model is created only after it successfully classifies 1000 classes contained in the Imagenet Database.A totally 1000 class of images exist in the Imagenet database and they are divided up into 3 folders: a training set with 1.2 crore images, a validation folder with 50,000 images and a test folder with 150,000 images.

VGG19 is a special type of convolution neural network that is trained on the Imagenet database and produces very good results.VGG19 stands for Visual Geometry Group.Below is the figure [Fig 1] which illustrates the layers of VGG19 model.

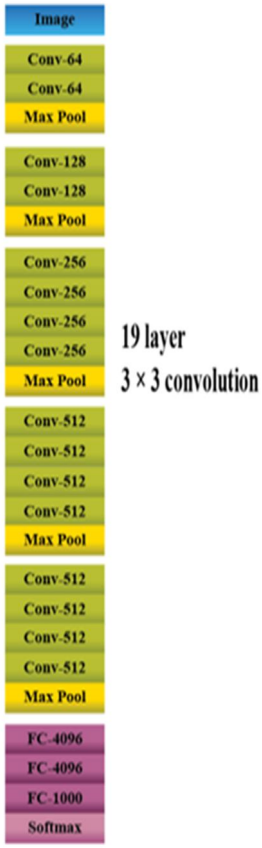


Fig.1. VGG-19 Architecture

VGG-19 has around 19 layers, and the images are passed through all 19 layers of CNN before they can be classified accordingly. Each layer in VGG-19 contains multiple 3x3 filters that can help deduce meaningful information from an image such as edge, lines etc.

Layers in the VGG-19 can be divided into three fully connected layers, five pooling layers, 16 convolution layers and one softmax layer. The last layer is the softmax layer, from which we get the classified image.All the layers in VGG-19 can be broadly classified in to the following three categories:

- *Convolution Layer:* In this layer, small 3x3 filters are applied to the input image that determine the image's key features, like edge, line, and so on. The filters in this layer identify features based on changes in intensity. As a result, this layer results in feature maps.
- *Pooling Layer:* Pooling is an important step in reducing the size of the feature maps generated by convolutional layers. The pooling layer reduces the issue of overfitting. In VGG-19, a special type of pooling is carried out referred to as maximum pooling.Max pooling resultsinfeature maps with the most relevant data.
- *Fully Connected Layer:* While the 2 layers mentioned above provide the important features, this layer is vital in classifying the image according to a specific class it belongs to.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

1) *Steps involved in VGG-19:*

a) It is recommended to provide an input image of 224*224 to this pre-trained model.
b) To prepare the images mean RGB value is subtracted from each pixel, which is calculated over the entire set of training images.
c) In order to derive the feature maps representing the most prominent features (such as edge, line, etc. a 3*3 sized filter is used.
d) Max pooling is used to reduce the size of feature maps.
e) Following that, an activation function RELU is used in order to create an output from the weighted sum of inputs. This RELU function helps in classifying the images more precisely.
f) There are then three fully connected layers, followed by a final layer called softmax.
g) The output from the final layer is the classified image.

2) *Advantages*

a) It is easy for training the image dataset using VGG-19.
b) Gives accurate results for image classification problem.

3) *Disadvantages*

a) It requires more memory as VGG-19 is a deep cnn.

B. *Inceptionv3 pre-trained model*

InceptionV3 model is a 42-layer, pre-trained model used to classify images. Out of the several variants of the Inception model, Inceptionv3 provides the most accurate classifications. There are several layers in it that carry out various operations like convolutions, average pooling, and max pooling. It is estimated that the Inception model is 78.1% accurate on Imagenet dataset. Below is the block diagram [Fig 2] which depicts InceptionV3 model.

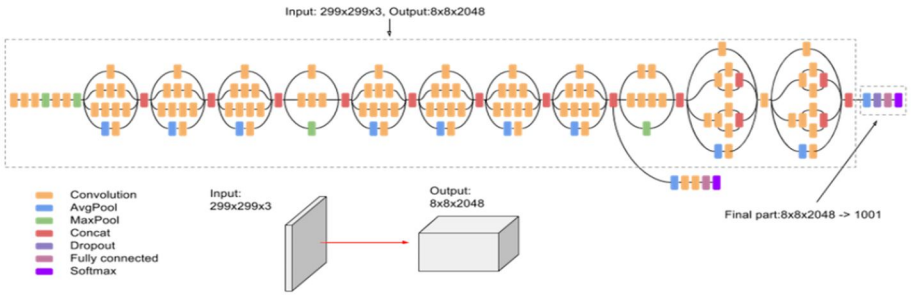


Fig.2. Inception model architecture

1) *Operations in Inception Model*

a) *Convolution:* Convolution operation is performed in order to merge information. The convolution operation involves applying kernels to each pixel in order to transform the image.
b) *Pooling:* By pooling, the dimensions of the feature maps can be reduced.

2) *Inception Model Architecture*

a) *Factorized Convolutions:* Ultimately, the network should have fewer parameters but computationally efficient.. The idea of a inception network is to reduce the number of parameters without sacrificing efficiency.
b) *Smaller Convolutions:* The convolutions in other pre-trained models are usually more complex, which results in longer training times. The inception model on the other hand, substitutes the larger convolution for a smaller convolution, thereby reducing training time.
c) *Asymmetric Convolutions:* Instead of using 3x3 convolutions, the Inception network uses 3*1 followed by 1*3 convolutions. .The network's main objective is to reduce the number of parameters by using the two filters discussed above.
d) *Auxiliary Classifier:* As part of inceptionv3, auxiliary classifiers are used to extend the network's depth. The classifiers act as regularizers in the network.
e) Grid size reductions: This technique is applied mostly in pooling stage. This is done to reduce the number of feature maps resulting from the max pooling operation.

3) Advantages

a) Efficiency is high.
b) Special classifiers are used in pre trained model as regularizers.
c) Deep network with around 42 layers.
d) Cost is low.

4) Disadvantages

a) Complexity is high.
b) It also leads to loss of information.

C. Resnet Pre-trained model

The ResNet acronym stands for Residual Network. There are around 100 and 1,000 layers in a ResNet. Layers are being added to the network, which will raise accuracy and performance automatically. There are several tasks associated with computer vision that require the ResNet model.

1) *Working of ResNet Network:* Adding layers will always result in an exploding gradient problem, but the residual network introduces the concept of skip connections. Stacks of Residual blocks form a ResNet. We have "skipped connections", which makes up the majority of ResNet. With this skipped connection approach, it will bypass a few layers of training and connect directly to the output.
2) *Architecture:* ResNet is constructed out of 34 layers. Upon capture, images are passed through a 7*7 filter with a stride of 2 and 64 channels.In addition to the three convolution layers in the ResNet model, there is a separate block, which has 3 convolution layers as well.Below is the block diagram [Fig 3] which depicts the ResNet architecture.

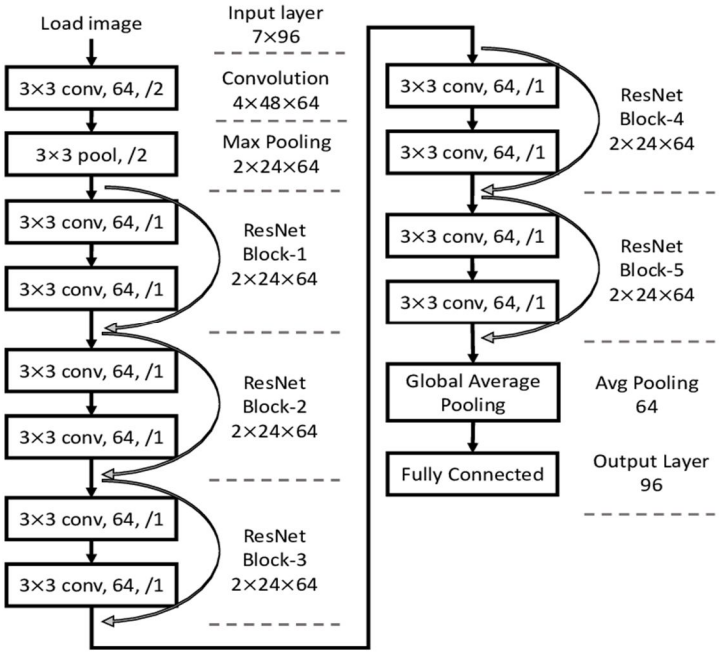


Fig.3. ResNet Architecture

3) Operations Involved in ResNet Model

Convolution and pooling are the two different operations carried out in ResNet model.

a) *Convolution:* In order to extract feature maps from images, convolution is applied. The technique uses small kernels over images to extract their features.
b) *Pooling:* Mostly max and global average pooling is carried out in ResNet architecture.Global average pooling is the process of replacing the fully connected layers so that only one feature map is generated as a result.Maximum pooling reduces dimensions by finding the maximum value and replacing all the values in the batch with it.

4) Advantages

a) Training is easier because of deep layers.

b) Helps in solving the vanishing gradient problem.

5) Disadvantages

a) Layers are numerous, hence it increases complexity.

b) Cost is high.

D. Densenet pre-trained model

Densenet is a convolution neural network containing deep layers as well. Densenet architecture focus on shortening the connections between the layers. This structure differs from other architectures because every layer is interconnected in the densenet model. Similar to other pre-trained models, this densenet model includes convolution and pooling.

This algorithm is composed of a convolution block, a dense block, and another dense block followed by a transition layer, followed by another dense block and a classification layer.

1) Components of Densenet Architecture:

a) Connectivity: The feature maps generated after convolution are concatenated into one feature map and then used as input. This way of concatenation is an easy approach for implementation purpose.

b) Dense Blocks: These dense blocks change the number of filters to reduce the dimension of feature maps. In addition to these dense blocks, there is a transition layer that, in turn, reduces the number of channels to half.

c) Growth Rate: By going through the layers, the feature maps also increase in size. This increase in size is avoided by introducing a parameter which controls the amount of information added to the network.

d) Bottleneck Layers: A 1x1 convolution layer is added before every convolution layer to maximize the efficiency of the pre-trained model. Below is the block diagram [Fig 4] of Densenet architecture.

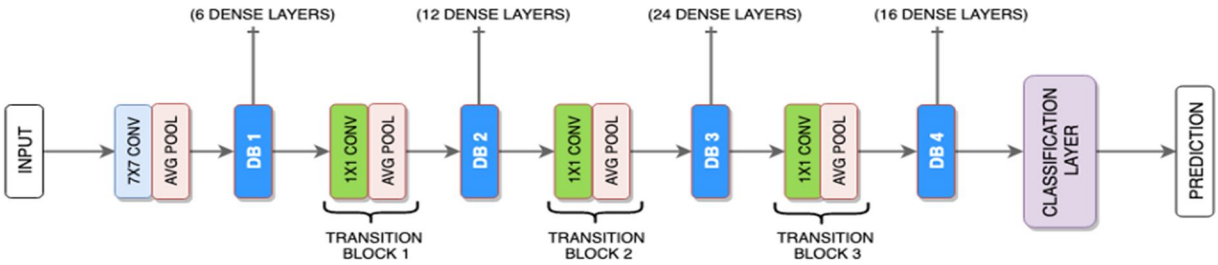


Fig.4. Densenet architecture

2) Advantages

a) Deep neural network.

b) Reduction in number of parameters

c) Eliminate Vanishing gradient problem.

3) Disadvantages

a) Cost is high

b) Highly complex

E. Xception Pre- trained Model

Xception is another example of a deep neural pre-trained model. In comparison to other pre-trained models described above, the Xception architecture resulted in more accuracy. The Xception model is composed of 71 layers .This is an enhanced version of the Inception model.

The Xception pre-trained model performs a depth-wise convolution instead of a normal convolution operation, as in other pre-trained models. After the depth wise convolution comes the point wise convolution. This results in a lighter model.In some cases, point wise convolution is followed by depth wise convolution.Below is the diagram [Fig 5] which depicts the block diagram of Xception pre trained model.

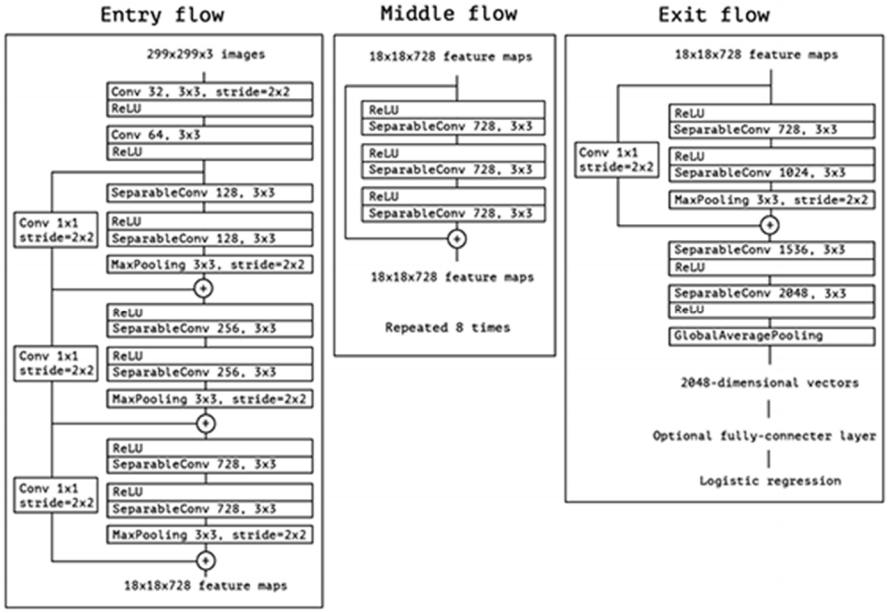


Fig.5. Xception Architecture

1) Advantages

a) Deep convolution network

b) Training is faster

c) Lighter model

2) Disadvantages

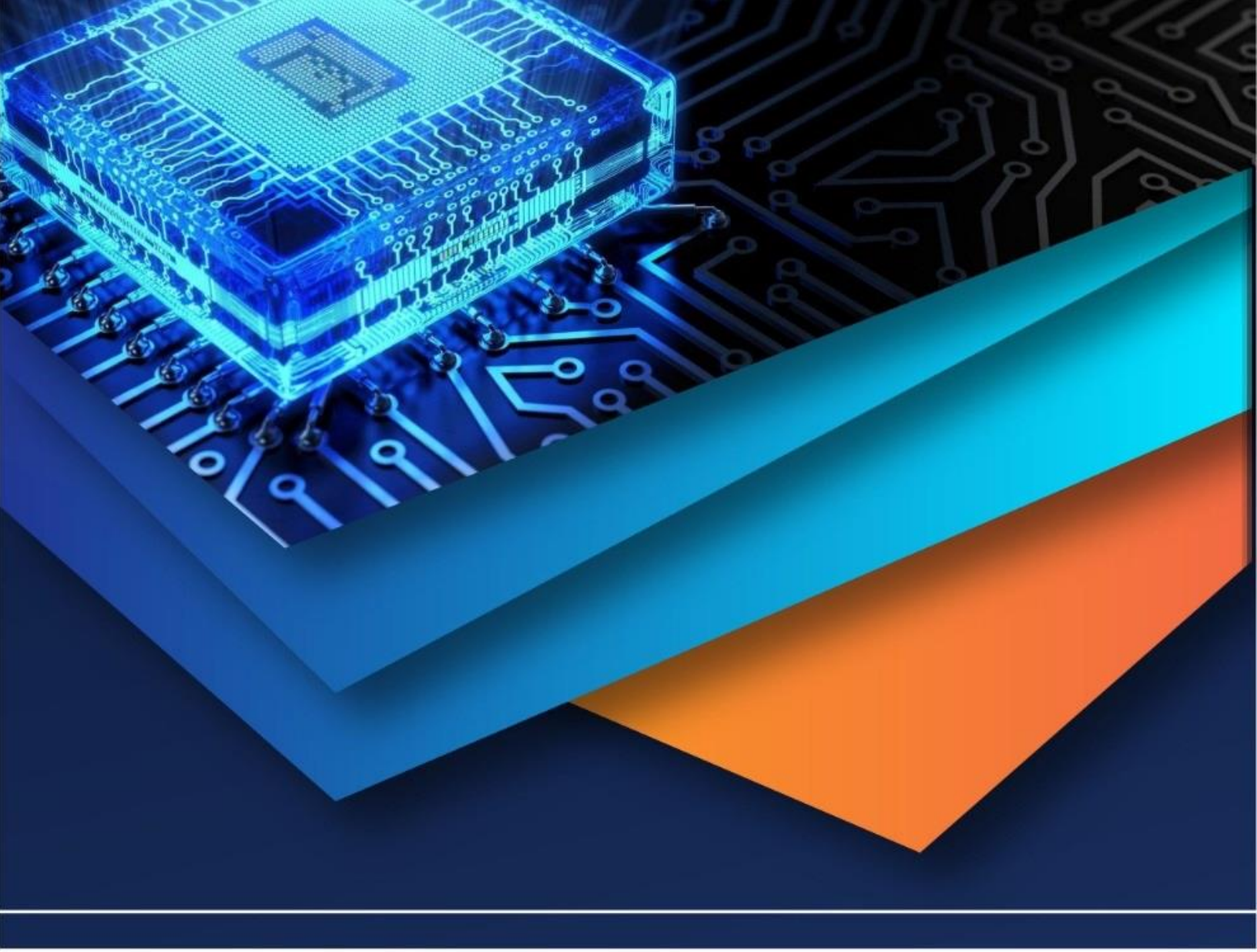
a) Highly complex

VI. CONCLUSION

In order to classify images with more accuracy, different pre-trained models are employed. Depending on the size of the dataset and machine capacity, a specific model is chosen and images are trained to maximize the accuracy of the classification.

REFERENCES

[1] Nick Babich, “What is computer vision and how does it work? An Introduction” Available at: https://xd.adobe.com/ideas/principles/emerging-technology/what-is-computer-vision-how-does-it-work/ accessed on 28-Jul-2020.
[2] Austin, “How computers see images”. Available at: https://realpython.com/lessons/how-computers-see-images/
[3] W. Sae-Lim, W. Wettayaprasit and P. Aiyarak, "Convolutional Neural Networks Using MobileNet for Skin Lesion Classification," 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2019, pp. 242-247.
[4] Illija Mihajlovic,”Everything you ever wanted to know about computer vision”, Available at: https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer- vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e accessed on April 26, 2019.
[5] ”Understanding the vgg19 architecture”, Available at: https://iq.opengenus.org/vgg19-architecture/
[6] A Convolutional Neural Network for Lentigo Diagnosis - Scientific Figure on Research Gate. Available at: https://www.researchgate.net/figure/Architecture-of-the-InceptionV3-model_fig4_342431074 accessed on 19 Dec, 2021.
[7] Sik-Ho Tsang,”Review.Inceptionv3 -1st runner up (Image classification)in ILSVRC”. Available at: https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c
[8] Short-Term Load Forecasting based on ResNet and LSTM - Scientific Figure on Research Gate. Available at: https://www.researchgate.net/figure/The-structure-of-ResNet-12_fig1_329954455 accessed on 19 Dec, 2021.
[9] ”Architecture of DenseNet-121“, Available at: https://iq.opengenus.org/architecture-of-densenet121/
[10] Ziliang Zhong, Muhang Zheng, Huafeng Mai, Jianan Zhao and Xinyi Liu, "Cancer, image classification based on DenseNet model",DOI:10.1088/1742-6596/1651/1/012143,published on: 23 Nov 2020.
[11] C. Wang et al., "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," in IEEE Access, vol. 7, pp.146533-146541, 2019.
[12] Devvi Sarwinda, Radifa Hilya Paradisa, Alhadi Bustamam and Pinkie Anggia, “Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer, Procedia” Computer Science, Volume 179, 2021, ISSN 1877-0509, 2021.
[13] Bansal, Monika & Kumar, Munish & Sachdeva, Monika & Mittal, Ajay, "Transfer learning for image classification using VGG19: Caltech-101 image data set.” Journal of Ambient Intelligence and Humanized Computing. 10.1007/s12652-021-03488-z, 2021.
[14] L2M Xception: an improved Xception network for classification of peach diseases, VL - 17, DO - 10.1186/s13007-021-00736-3.



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)