International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Deep Learning Model for Image Classification Using Convolutional Neural Network

Shaik Parvez

Department of Computer Science Engineering, Gayatri Vidya Parishad College of Engineering, Jawaharlal Nehru Technological University Kakinada (JNTUK), Visakhapatnam, India
shaikparvez977[at]gmail.com

Abstract: ***In many signal and image applications, deep learning has emerged as a crucial field of machine learning. It is a complicated process that depends on various elements. The application of one of the powerful deep learning algorithms, the Convolutional Neural Network (CNN), is picture categorization. The main goal of the work presented in this study is to detect translation, scaling, and other types of distorted and invariant image distortion with the help of deep convolution neural networks. This CNN model will be trained to detect different Hotel Room images and classify them based on trained data. A Flask application will be built to detect the image uploaded by the customer. In the background of this Flask application, the hotel room image classification model (CNN) will be running to detect the images. This Flask Server will be deployed on AWS (Amazon Web Services) EC2 (Elastic Compute Cloud) Machine so that users from all over the world can access it using either the Internet Protocol (IP) address of the machine or by using the DNS (Domain Name System) name assigned to that corresponding IP address. The researcher used the VGG16 (Visual Geometry Group) model of CNN architecture to build the model. For visualizing the accuracy of the built model, the Matplotlib library built on NumPy arrays is used. The experimental result analysis based on the graphical representation and quality metrics shows that the CNN algorithm provides moderately better classification accuracy for all tested datasets.***

Keywords: Artificial Intelligence**,** Image Classification, Deep Learning, Convolutional Neural Network Algorithm, Keras, Visual Geometry Group model

1. Introduction

Image classification is a type of supervised learning issue where a collection of target classes (things to identify in photographs) is defined. Based on the tagged sample photos, a model is trained to recognize the target classes. The development of the Convolutional Neural Network (CNN), which can extract higher-level representations of the visual material, was a significant advancement in the process of creating models for image classification. A CNN essentially learns how to extract these options and determine what item they contain by using the image's raw component data as input rather than preprocessing the data to produce options like textures and forms.

Machines are trained using training data that has been labelled accurately, and then they anticipate the outcome using supervised learning, a form of machine learning concept. The labelled data refers to some input data for which the suitable output label has been applied. The supervisor instructs the computers to accurately forecast the outcome using the data (training data) that is supplied to them. It makes use of a similar idea to how a pupil learns while being supervised by a teacher. Supervised learning is a method of providing the machine learning model with the appropriate input data, input knowledge, and output data. The aim of supervised learning is to find a mapping function to connect the input variable (x) with the output variable (y). In the actual world, supervised learning is widely used for tasks like risk assessment and image categorization, spam filtering, fraud detection, etc. In supervised learning, the models are trained using labelled datasets so that they may learn about various types of data. After the training phase is finished, the model is evaluated using test data, which is a subset of the training set, and makes a prediction of the outcome.

Deep Learning is usually referred to as a subset of Machine Learning, which is a subset of Artificial Intelligence. Artificial intelligence (AI) refers to methods that give computers the ability to imitate human behaviour. On the other hand, Deep Learning is a branch of Machine Learning that draws inspiration from the structure of the human brain. Algorithms in deep learning typically use a predetermined logical structure to examine the data and draw conclusions that are comparable to those reached by humans. Neural networks, a multi-layered structure of algorithms, are used by deep learning to do this.The different layers of neural networks can also be viewed as a finely tuned filter that increases the likelihood of detecting and delivering accurate results. Equally, the human brain functions. The brain attempts to make connections between new information and well-known objects whenever we get it. Deep neural networks employ the same idea. We can carry out several tasks using neural networks, including regression, classification, and agglomeration. With neural networks, we may classify or group the unlabeled data based on how closely the samples in the training data resemble each other. To categorize the samples in this dataset into several classes in the classification example, we will train the network on a labelled dataset.

2. Literature Review

Object identification and image analysis both heavily rely on picture classification. There have been many different image classification methods put forth so far. In an effort to determine the optimum image categorization method, numerous studies have been undertaken. Over the past few decades, conventional methods have been regularly altered to produce the most accurate results possible, and new picture categorization techniques have also been created.

Licensed Under Creative Commons Attribution CC BY

The current focus of this project is combining the desired aspects of various strategies to boost a typical dataset's efficiency.

Abhinav N Patil [1] from Vishwakarma University, India, performed research on classification of "imagery data using custom neural network with the architecture of Convolution Neural Networks and Keras API. His aim was to classify the images based on the built model (using Convolutional Neural Network Algorithm) with higher accuracy. Results were fluctuated a bit but according to the average, the accuracy was well around 90-95% percent with a layer filter of 256. More powerful hardware could have achieved even higher results and with much extended dataset for more categories than just two for training."

Yechuri Sandeep [2] from Satyabhama Institute of Science and Technology, India, done research on deep learning algorithms for image classification. "Created an image classification model using two methods. The first one was a traditional pattern recognition model. He extracted some human-crafted features like color and Dense-SIFT, represented images using bag of words model, and then trained Support Vector Machines(s) classifiers. For the second approach, he used Deep Convolutional Neural Networks (CNN) to learn features of images and trained Backpropagation (BP) Neural Networks and SVMs for classification."

Shanya Sanjay Verma, Dr. Sureshkumar N [3], used Convolutional Neural Network algorithm to perform image processing of crop yield classification." They also used data science techniques for collecting the data and shredding it. It was highly aimed toward Image processing hence we will be using the Image Processing frameworks present along with some Data Science frameworks. The created machine can identify crop and the most important part is that using advanced Image Processing techniques we were able to achieve this small initial database of images."

Leslie D [4], from The Alan Turing Institute, conducted experimental research on "Understanding bias in facial recognition technologies" that stated, "CNNs break down a digital image's two-dimensional array of pixel values into smaller parts, but instead of sliding rectangular shapes across the image looking for matches, they zoom in on particular patches of the image using smaller grids of pixels values called kernels."

3. Methodology

The primary goal of this study is to develop a more accurate picture categorization model utilizing convolutional neural networks (CNN). In this research, input will be an image provided by the user/ customer. The user/ customer will upload the image and has to click on submit button to classify that image. The image uploaded will be undergone through various pre-processing techniques like padding, fitting, etc. Based on the trained model, the pixel data and RGB data will be authorized, and the image will be classified among the categories. If the user does not provide any image and clicks on submit button, then an error pop-up message will be initiated. All the processing will be done using the trained model which is stored over the cloud. Different outputs will be generated depending on what module the users take or what input they will give. Based on the type of image provided, the model will categorize it. For example, Bedroom, Bathroom, Gym, Swimming Pool, etc. The entire process involved is represented in the form of a control flow diagram in the below fig-1.

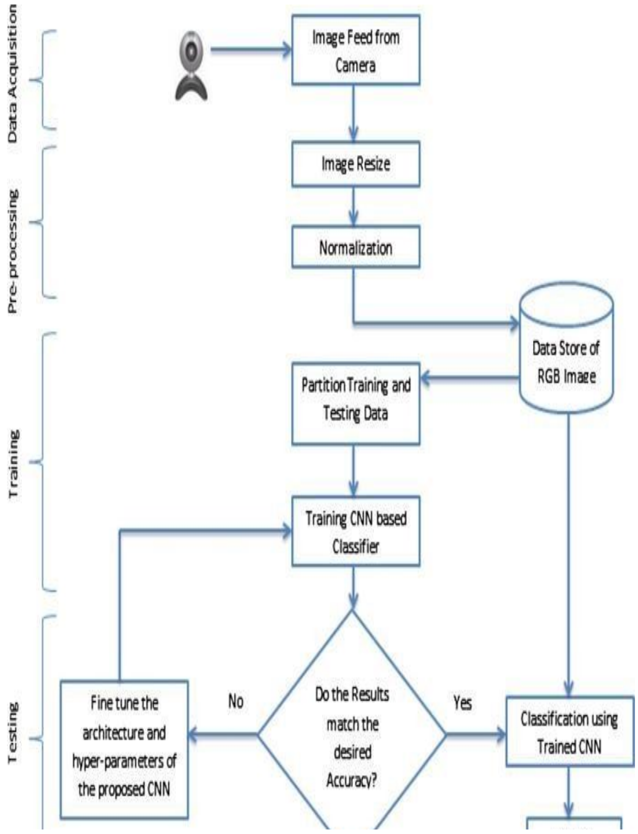

Figure 1: Control flow diagram

Three crucial steps make up the overall image classification process: image processing, image feature extraction, and classifier.

3.1 Image Processing

Image processing entails digitizing an image and applying various techniques to it in order to extract some useful information. When implementing a set of prescribed signal processing processes, the image processing system normally interprets all images as 2D signals. The steps performed in this research are illustrated in the below fig-2

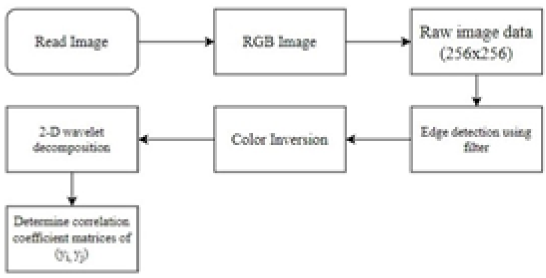

Figure 2

Licensed Under Creative Commons Attribution CC BY

3.2 Image Feature Extraction

Computer vision and image processing both use the concept of feature extraction. It describes the procedure of gathering image data with the aid of a computer and determining if the points on a specific image are included in image feature extraction. The division of the image's points into several subsets is the aim of feature extraction. Frequently, these subsets consist of an area, a continuous curve, or single points.

“The time-frequency composite weighting algorithm for multi-frame blurred images is a frequency-domain and time-domain weighting simultaneous processing algorithm based on blurred image data. Based on the weighted characteristic of the algorithm and the feature extraction of the target image in the time domain and frequency domain, the depth extraction technique is based on the time-frequency composite weighting of the night image to extract the target information from the depth image” [5].

The time-frequency weighted feature extraction approach involves the following primary steps:

Step 1: Create a time-frequency weighted signal model for a number of blurry photos.

Step 2: After translating the one-dimensional function of the time scale and the time shift to the two-dimensional function of the time scale and the time shift, transform the continuous nighttime image of the image using a time-frequency composite weighted with a square-integrable function.

Step 3: Create a time-frequency composite weighted signal form

Step 4: The multi-thread fuzzy image's time-frequency weighted signal's frequency modulation law is a hyperbolic function.

Step 5: Apply the time-frequency weighting to the image using the multi-detector fuzzy image weighted signal image transformation formula (the definition of the image transformation is similar to the formula).

An image time-frequency weighted image signal will be the output in the end. Therefore, the time-frequency composite weighting algorithm can better realise this image feature extraction technique than the conventional time-domain.

3.3 Classifier

A classifier is then required to categorize the feature vectors once the feature vectors have been extracted using the image and the image has been characterized as a vector of static length. A input layer, activation layer, convolution layer, complete connection layer, pool layer and final output layer are the layers that make up a standard convolution neural network from input to output. Layer by layer, input data is transferred from the convolutional layer, which establishes the connections among various computational neural nodes, while the continuous convolution-pool structure decodes, infers, transfers the original data's feature signals to the feature space of the hidden layer. The following entire connection layer identifies and produces using the features that were extracted. Steps involved in the process of classification of raw data is illustrated in fig-3.

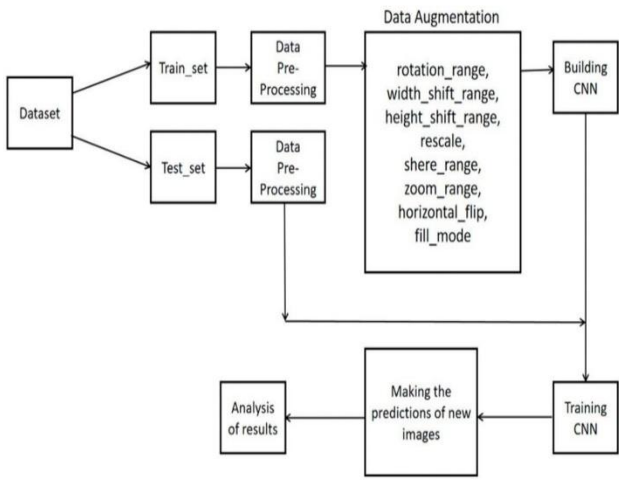

Figure 3: Classification of raw data

4. Experimental Setup

The images of any type of hotel room can be trained on this network. However, the machine must meet certain hardware and software requirements.

H/W Requirements (Application Specific Hardware):
- Installed libraries – PIL, NumPy, pandas, keras, Matplotlib, TensorFlow, h5py
- Processor i3
- UNIX based OS
- 4Gb RAM
- Required environments: Ipython, python 2.7
- 2Gb graphic card

S/W Requirements (Application Specific Software):
- Python-compatible environment (python2.7+) for training the convolution network.
- K-20 GPU enabled EC2

5. Convolutional Neural Network

A deep learning approach used for object recognition is a convolutional neural network. The idea behind CNN is to teach the network how people learn new things. A massive dataset is used to train a CNN. A unique method is used by each layer of the CNN to extract features from the image. Since all the layers are interconnected, all the extracted features can be integrated at the end. To distinguish between human and non-human faces, CNN will also undergo training using non-human faces. To calculate the characteristics, CNN uses iterative calculations.

Licensed Under Creative Commons Attribution CC BY

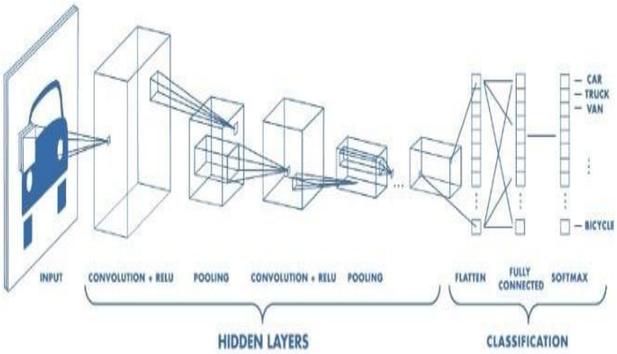

Figure 4: Layers of Convolutional Neutral Network

6. Visual Geometry Group-16 Model

Visual Geometry Group-16is a convolution neural network (CNN) architecture which was proposed in the ImageNet competition in 2014.This algorithm can correctly identify 1000 photos from 1000 different categories by detecting objects in each image. It is a well-liked approach for classifying images that functions well with transfer learning.

It is widely recognised as one of the greatest vision model architectures ever made. The key distinction of VGG16 is that it prioritised having Maxpool layers of 2x2 filters with a stride 2 in addition to convolution layers of 3x3 filters with a stride 1 and always used the same padding.The VGG16 working and design is visualized below in fig-5.

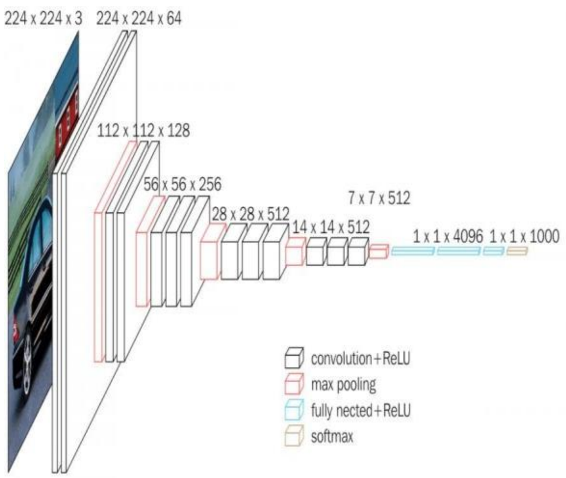

Figure 5: VGG16 Architecture

To build the VGG16 model, various number of packages have been used. Below code shows the packages used:

```
from keras.layers import Dense, Flatten
from keras.models import Model
from keras.applications.vgg16 import VGG16
from keras.applications.vgg16 import preprocess_input
from keras.preprocessing import image
from keras.preprocessing.image import ImageDataGenerator
from keras.models import Sequential
from keras.preprocessing.image import load_img
from keras.preprocessing.image import img_to_array
#from keras.applications.vgg16 import decode_predictions
import numpy as np
from glob import glob
import matplotlib.pyplot as plt
```

Figure 6

7. Testing

Based on the dataset, the model is evaluated using model.compile() and model.fit_generator() function, and fig-7 shows the accuracy of the model on test data and the accuracy of the model on training data.

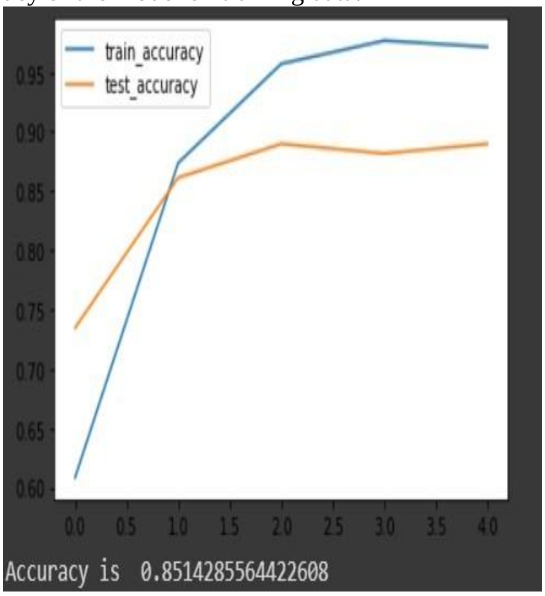

Figure 7: Graphical representation of accuracy

The model is trained with required epochs and in the picture, we can see the loss and validation loss pair and the accuracy of training data and validation data accuracy. Below fig-8 shows the detailed information on training which involves the total number of epochs, loss, train_accuracy, and val_accuracy.

```
Epoch 1/5
18/18 [==============================] - 512s 28s/step - loss: 0.9720 - accuracy: 0.6085 - val_loss: 0.3414 - val_accuracy: 0.7347
Epoch 2/5
18/18 [==============================] - 418s 23s/step - loss: 0.3150 - accuracy: 0.8737 - val_loss: 0.3421 - val_accuracy: 0.8612
Epoch 3/5
18/18 [==============================] - 425s 24s/step - loss: 0.1692 - accuracy: 0.9573 - val_loss: 0.0982 - val_accuracy: 0.8898
Epoch 4/5
18/18 [==============================] - 423s 24s/step - loss: 0.1195 - accuracy: 0.9769 - val_loss: 0.3146 - val_accuracy: 0.8816
Epoch 5/5
18/18 [==============================] - 423s 24s/step - loss: 0.1106 - accuracy: 0.9715 - val_loss: 0.4453 - val_accuracy: 0.8898
```

Figure 8

Licensed Under Creative Commons Attribution CC BY

8. Results

After implementing the models, they are deployed on the AWS EC2 engine with the help of flask. Below fig-9 shows the home screen when a user enters the URL. The URL can be either IP Address or Domain Name for that corresponding IP the interface of the app in which the user has to upload the image.

Example-1: Fig-10 shows the output when user uploads a Bedroom image on the home screen. When the user uploads a Bedroom image then the model at the back will try to detect the uploaded image as Bedroom and then redirects the image to the following path: /var/www/html/uploads/Bedroom_Images

Figure 10

Example-2: Fig-11 shows the output the output when user uploads a Swimming Pool image on the home screen. When the user uploads a Swimming Pool image then the model at the back will try to detect the uploaded image as Swimming Pool and then redirects the image to the/var/www/html/uploads/SwimmingPool_Images.

Figure 11

Example-3: Fig-12 shows the output when user uploads an image of the gym in the home screen. When the user uploads the image of the gym then the model at the back will try to detect the uploaded image as gym and then redirects the image to the following path: /var/www/html/uploads/Gym_Images

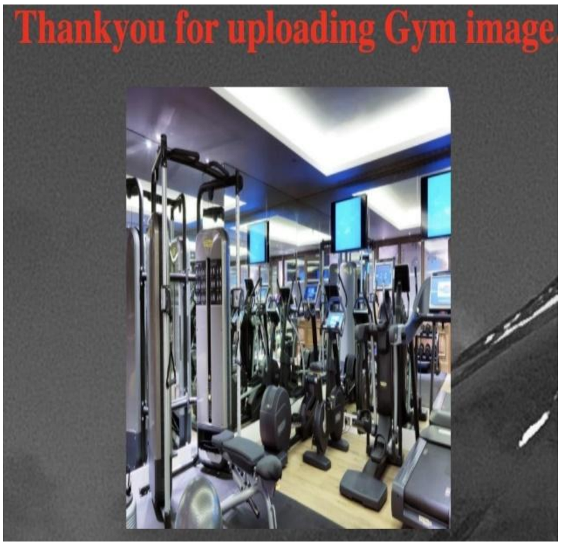

Figure 12

9. Conclusion

In this project, we used Convolutional Neural Networks (CNN) for image classification using images from hotel room data sets. Using CNN, this set of data was used for both training and testing. It provides an accuracy rate of 98% for the training dataset and 85% for the test dataset using the Vgg16 model. Images used for the training purpose are small and RGB images. The processing of these photographs takes a significantly long amount of time computationally when compared to typical JPEG images. Results for image classification will be more accurate if the model is built with more layers and trained on more picture data using GPU clusters.

Licensed Under Creative Commons Attribution CC BY

The future enhancement of classifying images of large sizes is very useful for the image segmentation process. The main aim of this project is that whenever a user uploads an image to give a review to the hotel, then one person should sit at the back to arrange these images into their corresponding folders. So, this is a hectic work for the person. Our project focuses to make this automation. So, when the user uploads an image on the website then the VGG16 model which is attached at the backend will detect the image and redirects the image to its corresponding folder. And this Flask server is running on an AWS EC2 machine, user can access this website from anywhere in the world by using either the IP address of the machine or by using the DNS name assigned to that corresponding IP address.

References

[1] Patil, Abhinav, Image Recognition using Machine Learning (February 1, 2021). Available at SSRN: https://ssrn.com/abstract=3835625 or http://dx.doi.org/10.2139/ssrn.3835625.
[2] Yechuri Sandeep," Deep Convolutional Neural Networks (CNN) to learn features of images and Backpropagation (BP) Neural Networks and SVMs for classification.," Satyabhama Institute of Science and Technology, India, 2014.
[3] Shanya Sanjay Verma, Dr. Sureshkumar N, "Image Processing for Crop Yield Classification", International Journal of Science and Research (IJSR), Volume 8 Issue 12, December 2019, pp. 1768-1771, https://www.ijsr.net/get_abstract.php?paper_id=ART20203802
[4] Leslie, David, Understanding Bias in Facial Recognition Technologies (September 26, 2020). Leslie, D. (2020). Understanding bias in facial recognition technologies: an explainer. The Alan Turing Institute. https://doi.org/10.5281/zenodo.4050457, Available at SSRN: https://ssrn.com/abstract=3705658 or http://dx.doi.org/10.2139/ssrn.3705658
[5] Xin, M., Wang, Y. Research on image classification model based on deep convolution neural network. J Image Video Proc. 2019, 40 (2019). https://doi.org/10.1186/s13640-019-0417-8

Licensed Under Creative Commons Attribution CC BY