IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Scene Understanding Of Nature Using CNN

Dr. P. Shanmugapriya , Tammana Karthikeya,V.Narasa Reddy

Associate professor, B.E graduate(IV year), B.E graduate(IV year)

Department of Computer science and engineering

SCSVMV, Kanchipuram, India

Abstract:

Our research's overarching goal is to comprehend an image by categorizing its properties and grouping them under the appropriate scenes. To achieve the same goal, They created a convolutional neural network model from a data set that contained 6000 trained and 3000 test photos. Ourselves then tested the model using images from various categories. Six classes comprise the dataset, providing a comprehensive picture of nature worldwide. The dataset is tested and categorized after our model has been trained using a succession of convolution layers, a pooling layer, and a fully connected layer to obtain an output containing photos sorted according to scene categories.

Keywords: .

Convolutional Neural Networks, Image Processing, Filters Max Pooling, Batch Normalization, Scene recognition, Padding, Drop out layer, Pooling layer, PLACES Dataset.

Introduction:

The goal of computer vision research has been to enable machines to have a vision similar to that of humans. This project's key component is scene understanding, which enables the definition of an object recognition context. Scene understanding is a key component of computer vision because it allows for the real-time perception, analysis, and development of interpretations of dynamic scenes that result in new insights.

It aims to make any image as comprehensible and discernible to computers as it is to people. Google Photos, intelligent robotics, human-computer interaction, autonomous driving, and sophisticated video monitoring are just a few areas where scene recognition has been widely applied. Essentially, it is an ordered perspective of the environment in real life, which includes a variety of surfaces and things.

Objectives:

The view is to address one of the most prevalent scene recognition issues. Approximately 6000 trained and 3000 tested images of various settings make up the PLACES dataset, which was useful for experimentation. This dataset is sufficiently comprehensive to train and test a model to provide the necessary output. A neural architecture for our machine to store and classify the photos as requested for the recognition of scenes in the dataset was created using the convolutional neural network. The obtained results strongly imply that the trained model could successfully recognize the nature scenes in the given data collection. Scene understanding is the process of observing, evaluating, and developing an interpretation of a 3D dynamic scene, frequently in real time.

Scope of project:

Understanding natural situations and recognizing and utilizing CNN, which can be used by various applications to complete their visualizations, are the goals of this work. on CNN called a convolutional neural network is made specifically to process pixel data and used for image recognition and processing. The word "convolution" refers to one of the network's most significant operations. The brain serves as the model for convolutional neural networks. Convolutional, pooling, and fully connected layers make up the typical architecture of a CNN. Compared to other image classification algorithms, CNN uses comparatively little pre-processing. The main

benefit of CNN over its forerunners is that it does so without human intervention, automatically identifying the key features. CNN's lack of encoding of object position and orientation is a drawback. being spatially insensitive to the input data.

Existing systems:

The quality of the input has a significant impact on recognition accuracy in the majority of the current systems. Images frequently touch or overlap. Scene recognition is assessed in the majority of the segmentation algorithms currently in use. Additionally, scene recognition varies from location to location. This calls for using statistical classifiers and artificial neural networks to extract data.

Literature survey:

Scene identification is an area that is quickly expanding and has attracted a lot of interest recently. It is a necessary stage for many applications, including robot navigation and map generation, among others. Convolution Neural Networks (CNN) and other deep learning techniques have been developed to achieve better scene representation. When comparing the density and diversity of image datasets, Zhou, Bolei, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva (2014) introduced a new scene-centric database called Places with over 7 million labeled images of scenes and proposed new methods. They demonstrated that Places is as dense as other scene datasets and has a greater level of diversity.[1] They develop new state-of-the-art findings on numerous scene-centric datasets using CNN, which they use to learn deep features for scene recognition tasks.
They were able to demonstrate variations in the internal representations of object-centric and scene centric networks by the visualization of the CNN layers' responses. IN multi-scale convolutional neural network (CNN) architectures, scale-induced dataset bias, and how to use Places and ImageNet to successfully blend scene-centric and object-centric knowledge in CNNs. As their past work with Hybrid-CNN demonstrated, adding ImageNet did not significantly improve their performance. As a result, they developed a different approach that took scale into consideration and saw considerable recognition gains. They discovered that ImageNet-CNNs and Places-CNNs operate in distinct scale ranges and that utilizing the same network for all scales introduces dataset bias, resulting in limited performance.[3] Their research revealed that the scale has a significant impact on recognition accuracy and that using straightforward but carefully considered multi-scale combinations of ImageNet-CNNs and Places-CNNs, the state-of-the-art recognition accuracy in SUN397 can be increased to 66.26% (and even 70.17% with more complex architectures, on par with human performance)Bogdan Kwolek (2014) introduced a deep neural network-based system for indoor place recognition that employs transfer learning to retrain VGG-F, a pre-trained convolutional neural network, in order to categorize locations on photos captured by a humanoid robot.[2]

To enhance performance For scene detection and visual domain adaptation, Guo-Sen Xie, Xu-Yao Zhang, Shuicheng Yan, and Cheng-Lin Liu (2015) proposed combining CNN with dictionary-based models (DA)[4].In particular, the mid-level local representation (MLR) and convolutional Fisher vector (CFV) representations are further developed based on the well-tuned CNN models (e.g., AlexNet and VGG Net). A class-mixture or a class-specific part dictionary is produced in MLR using an effective two-stage clustering algorithm, which involves weighted spatial and feature space spectral clustering on the parts of a single picture followed by clustering all representative parts of all images. The midlevel representation is then created using the part dictionary in conjunction with the multiscale picture inputs. In CFV, Fisher vectors are produced based on the final convolutional layer of CNN using a multiscale and scale-proportional Gaussian mixture model training technique. Modern performance on scene identification and DA problems can be attained by combining the complementary data of MLR, CFV, and the CNN features of the fully connected layer. Their proposed hybrid representation (derived from a VGG net trained on ImageNet) contains an intriguing discovery in that it significantly complements Google Net and/or VGG-11.In 2016, Pengjie Tang, Hanli Wang, and Sam Kwong introduced a multi-stage feature fusion method based on Google (G-MS2F). The three outputs that correspond to the three components of the proposed model are used to construct the final choice for scene recognition using the product rule.[5] The experimental findings show that the suggested paradigm works. superior to several modern CNN scene recognition models, and achieves recognition accuracy of 92.90%, 79.63%, and 64.06% on the benchmark scene recognition datasets Scene15, MIT67, and SUN397, respectively.[6]

Problem statement:

The enormous variability and ambiguity of nature from place to place. The variation and ambiguity in the natural world. The nature scenes will occasionally change to comprehend their features and provide scalability. The Source image or document that has degraded over time in quality. Scaling the photographs and using a filter to extract the key features and characteristics from the image makes it harder for us to get the great significance features and the dimensionality of the image when these images are in the JPEG format. Data is transformed from a high-dimensional space into a low-dimensional space using a process called "dimension reduction," which aims to keep the lowdimensional representation as close as possible to the intrinsic dimension of the original data.

Proposed system:

The displaying unlabelled photographs of nature in the suggested system. The project also has the benefit of using the convolutional neural network, which is having. The necessary libraries are first imported, and six categories—designated as 0 for buildings, 1 for the forest, 2 for glacier, 3 for mountain, 4 for sea, and 5 for street—are selected for scene classification.
The model is then trained using a training set of 6000 images, a testing set of 3000 validation images, and a total of 6000 images. A fully connected layer is added after a conv2D and pooling layer series to create a CNN model.

The conv2D layer with activation function 'Relu' and the pooling layer are sequentially fed the input image of size 128*128*1, which then enters the fully connected layer where the flatten and dense layers with activation function 'Softmax' are applied to classify images under the appropriate classes based on output from convolutional layers. The prediction set, which consists of 1531 images, is used to test the entire model. Images selected at random for prediction are successfully categorized into one of six categories, and the output indicates the scene to which each image belongs. For the aforementioned scene classification, an accuracy of 82% is attained.

HIGHLIGHTS:

Easier to process, more accurate, less complicated.

CHALLENGES:

Hard to analyse and Longer processing time for semi-structured data.

High complexity.

Modules:

There are two modules in this project.
They are:
1. Sequential module.
2. Keras module.

PRIMARY MODULE:

Module – 1:

Sequential module:

In this module, the extraction module from the TensorFlow and Keras, Then creates a Dense layer, Flatten layer, and Input Layer, reshaping, Batch Normalization, Dropout, Conv2D, and MaxPooling2D layers. Here firstly create the model and Add layers of Input with Flatten layer Hidden with a Dense layer and Output with a Dense layer. Then achieved Prescribe the output.

Module – 2:

Keras and Tensor Flow module:

Here TensorFlow is the low-level API so keras is used.

Functional model
These models help to build complex computation graphs. it is the extraction module from the

TensorFlow and Keras, Then create a Dense layer, flattens the layer and Input Layer, reshapes, Batch Normalization, Dropout, Conv2D, and MaxPooling2D layers. Add layers of Input with Flatten layer and Hidden with a Dense layer and Output with a Dense layer. create the model Prescribe the output.

Architecture:

First, CNN is quite useful and provides findings that are more accurate than those obtained using other methods. It has also attained visual recognition that is almost human-level. Our dataset includes 6000 training photos and 3000 testing images, each of which were categorised into a total of 6 categories. Comparatively speaking, convolutional neural networks require less pre-processing than other image classification algorithms. Through automated learning, the network develops the ability to optimize the filters or kernels.

A significant advantage of CNN is the way it deviates from past knowledge and human assistance in feature extraction.

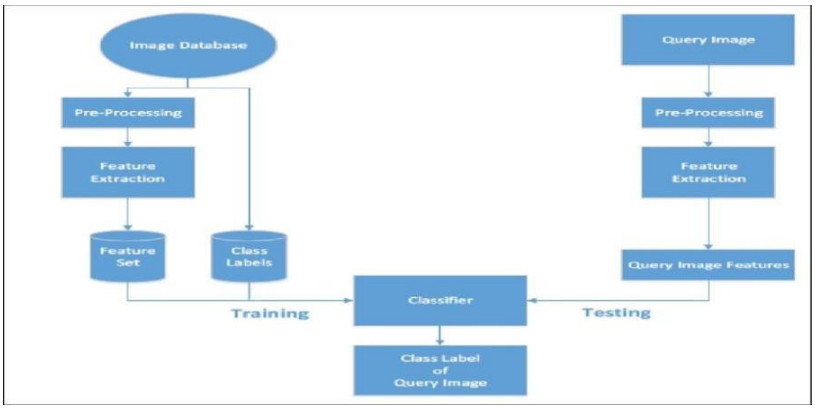

Figure 01: Sample Architecture

DATASET :

Here, the employed a dataset of scene-centric photos, with 6000 images for the training set and 3000 images for the testing set, both labelled with six categories. The dataset was then used to train the model. Later, 1531 photos were used for the label-based classification of other groups. The correctness of each category is also determined. Epochs are used for backpropagation and traversal of the decision tree and the random forest to get correct and high accuracy, and it gives a label to unlabelled images by classifying images under 6 categories and the accurate result was published. The system models 13 layers of dense so that it will easily identify the core features of the image.

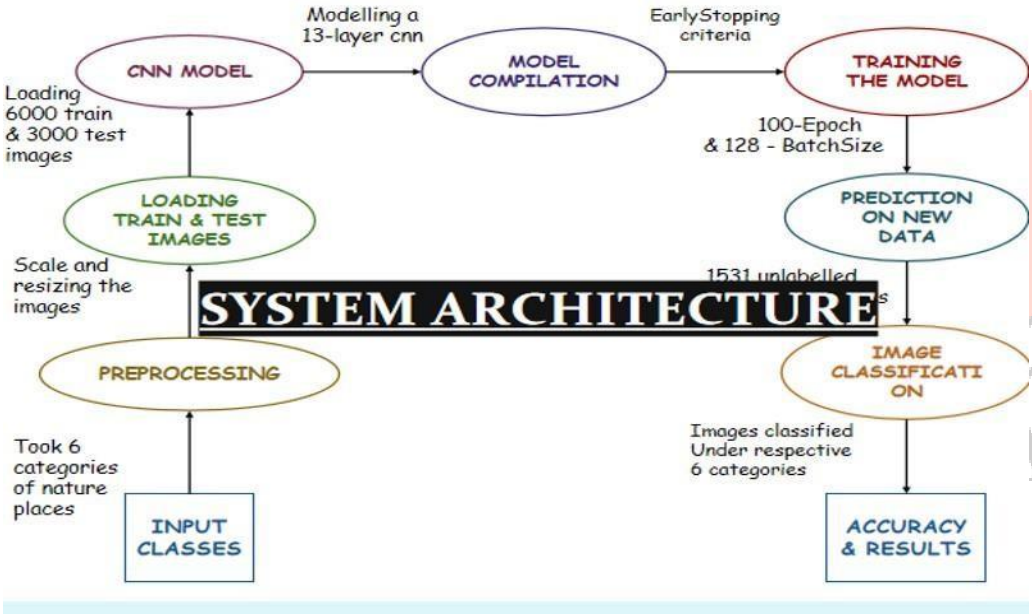

Figure 02: System Architecture

Process:

To enable a computer to see objects as a human would by employing a convolutional neural network. Here, the execution of the unstructured data is quick and effective thanks to the partially (semi)connected layer. Scene understanding is the perception, analysis, and development of an interpretation of a 3D dynamic scene seen through a network of sensors, frequently in real time. This method entails comparing signal data from sensors viewing the scene with models that people use to comprehend the scene. Scene comprehension is the process of adding semantics to and extracting it from sensor data that describes a scene. It extracts information about the physical world that is relevant to human operators in addition to the recognition of visual elements like corners, edges, and moving areas. It can accomplish detection, localization, recognition, and understanding at four levels of generic computer vision functionality. Convolutional is a mathematical term that describes joining two functions to create a third function. Convolutional layers are the layers on which filters are applied to the original image to extract the essential characteristics from it; they can be applied to other feature maps to produce a deep convolutional neural network.

Methodology:

In order to normalise the data, each pixel value was deducted from the average RGB value across all pixels in the image in this phase. The CNN's backpropagation training process, whose net input is the product of a node's inputs and weights, has been known to benefit from this normalisation.

A node can achieve the optimal gradient value, which is used for weight updating in the system, by having smaller net inputs. It is crucial to ensure that each feature has a comparable range in order to keep the gradients under control. Due to the concept of parameter sharing that these CNNs use, sharing of values will be difficult if the network inputs are not scaled to have comparable known ranges because different parts of the image may end up having values from entirely unrelated domains and ranges.

Implementation:

This describes Proposal work done by Ourselves have to import all the libraries or modules. First, Then deploy the smart contract and then use it.

Requirement specifications:

- Python Language.
- Google Colab.
- Google Drive.
- Solidity compiler.
- Image Folder.
- CNN.

Results:

The results are produced for the collection of prediction set, which consists of 1531 images, is used to test the entire model. Images selected at random for prediction are successfully categorized into one of six categories, photographs that are categorized according to the types of a scene they represent, such as mountains, streets, forests, seas, buildings, or glaciers and the output indicates the scene to which each image belongs. For the aforementioned scene classification, an accuracy of 82% is attained.

The project's outcomes are summarized as follows:

Figure 03: Output as Classified Images

Conclusion:

Scene Understanding is a style of seeing the physical world that involves multiple surfaces and objects that are organized in a meaningful way, used a model of convolutional neural networks to classify natural scenes with the highest degree of accuracy, and took a dataset of scene-centric photos made up of six categories representing various natural settings from around the world.

The results are produced for the collection of predicted photographs that are categorized according to the types of a scene they represent, such as mountains, streets, forests, seas, buildings, or glaciers.
The model's total accuracy was determined to be 82%.

Future scope:

The applications and challenges of this project are outlined in this study. Because CNN is so versatile in feature extraction, Because CNN is so versatile in feature extraction, it is the greatest option to replace the time- and work process of manual feature extraction. Images that combine two scenarios produce ambiguity, making it challenging for the model to categorize. As a result, it occasionally causes algorithms to fail. RGB graphics are utilized for training purposes. These photographs require more calculation time to process than typical images, on average. More layers in the model and more picture data used to train the network using clusters of GPUs will result in a more accurate classification of photos. Future improvements will concentrate on categorizing the coloured, largescale photographs in addition to adding more diverse category sceneries including waterfalls, and the sky. Additionally, try to identify and categorize the video's scenes.

References:

[1] Zhou, Bolei, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva(2014) introduced a new scene-centric database called Places with over 7 million labelled pictures of scenes and proposed new methods to compare the density and diversity of image datasets and showed that Places is as dense as other scene datasets and has more diversity.

[2] Piotr WozniakHadha AfrisalRigel Galindo EsparzaBogdan Kwolek(2014) proposed a deep neural network-based algorithm for indoor place recognition where it uses transfer learning to retrain VGGF, a pre-trained convolutional neural network to classify places on images acquired by a humanoid robot.

[3] *Luis Herranz*, *Shuqiang Jiang*, *Xiangyang Li* Computer Vision and Pattern Recognition (CVPR) Scene Recognition With CNNs: Objects, Scales and Dataset Bias

[4] To improve the performance Guo-Sen Xie; Xu-Yao Zhang; Shuicheng Yan; Cheng-Lin Liu(2015) proposed to combine CNN with dictionary-based models for scene recognition and visual domain adaptation (DA).

[5] PengjieTang, HanliWang, and SamKwong(2016) proposed Google Net based multi-stage feature fusion (G-MS2F). The product rule is used to generate the final decision for scene recognition from the three outputs corresponding to the three parts of the proposed model.

[6] SarfarazMasood[1]UmerAhsan[2]FatimaMunawwar[3]Danish RazaRizvi Mumtaz Ahmed Scene Recognition from Image Using Convolutional Neural Network This work is an attempt to solve one of the most common problems of scene recognition. The dataset called PLACES2 was used for the purpose of experimentation which contains nearly 7 million labeled pictures of various scenes which is a sufficiently exhaustive dataset to train and test a model to get the desired output.