Transfer Learning based Natural Scene Classification for Scene Understanding by Intelligent Machines

Ranjini Surendran[1], J. Anitha[1], A. Angelopoulou[2], E. Kapetanios[3], T. Chausalet[2] and Jude Hemanth D[1,*]

[1]Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India
[2]School of Computer Science and Engineering, University of Westminster, London, UK
[3]School of Physics, Engineering & Computer Science, University of Hertfordshire, UK
*Corresponding Author: judehemanth@karunya.edu

Abstract. Scene classification carry out an imperative accountability in the current emerging field of automation. Traditional classification methods endure with tedious processing techniques. With the advent of CNN and deep learning models have greatly accelerated the job of scene classification. In our paper we have considered an area of application where the deep learning can be used to assist in the civil and military applications and aid in navigation. Current image classifications concentrate on the various available labeled datasets of various images. This work concentrates on classification of few scenes that contain pictures of people and places that are affected in the areas of flood. This aims at assisting the rescue officials at the need of natural calamities, disasters, military attacks etc. Proposed work explains a classifying system which can categorize the small scene dataset using transfer learning approach. We collected the pictures of scenes from sites and created a small dataset with different flood affected activities. We have utilized transfer learning model, RESNET in our proposed work which showed an accuracy of 88.88% for ResNet50 and 91.04% for ResNet101 and endow with a faster and economical revelation for the application involved.

Keywords: Scene Classification, Image Classification, Deep Learning, Transfer Learning, Res Net

1 Introduction

Image classification [1] finds application in the field of automation, face detection [2], self-driving cars, robotics, aviation, civil and military applications. Many image classifications approaches classify the image based on single object or multi- object [3]. Scene classification has one of the most imperative roles in the field of scene understanding. Understanding a scene involves the steps akin to scene classification, object detection, object detection and localization and event recognition [4]. Scene consists of not merely objects, but preserves information regarding the relation and activity involved. Scenes carry a lot information regarding the content, relation and actions within the objects present. Identifying and analyzing the scene with the meanings preserved is significant task for computer vision applications.

Deep learning [5], [6] the subset of Artificial Intelligence has taken the classification to peaks. The availability of huge dataset and high computationally powerful ma-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

chines have led to the use of deep learning in classification. Many works are carried out in the field of image classification [7], [8], [9], [10], [11] with powerful deep learning models using publically available large indoor and outdoor datasets. Classifying images with deep learning finds many applications in our day to day life. With the inspiration from the deep architectures we tried to develop a classifying system for natural scenes. As we have seen natural disasters and calamities are affecting almost all countries in the world. We cannot not predict its effect and eliminate it, but can provide a helping hand for those affected areas. With this intension we developed a model that can classify few scenes representing the natural calamities. Here the disaster calamity of flood affected areas are showcased considering the people and vehicles if they are endangered. We have also considered scenes showing the rescue operations mainly with the aid of helicopters and boats to aid the rescuers to reach the calamity affected areas.

The dataset available were limited resource for our application. We created our own dataset consisting of few scenes. We had collected scenes of flood affected areas. We developed this model with the intention that this could serve humanity. Dataset is arranged considering scenes of people and vehicles that are drowned in flood and scenes of rescuing operation needed by boat or helicopter. Thus, our model could classify the scenes into four classes: Many people in flood, Vehicles in flood, Rescue by boat and Rescue by helicopter. These scenes could pave the way for the rescuing operations which would be hard enough for the people reaching out there. Deep models are pretrained on large dataset with high accuracy. They are trained on millions of images. Here we have used the concept of transfer learning for modelling our system. In transfer learning we can modify the structure of already learned models to our application. Since these models have already learned from the large image dataset collections, we can use it in our work where it can learn our small scene dataset.

2 Literature Review

Classifying an image is a simple job for humans, but when it comes with the machines it is not. Human can easily visualize a scene with its objects, relations these objects are carrying with each other and the different activities involved. The machines need to be trained with the features involved within the scenes. Traditionally these features were extracted manually using machine learning algorithms like SIFT, HoG and classification algorithms like SVM [12] and fuzzy classifier [13] carried out the classification.

Convolutional Neural Network (CNN) enhanced the capability of machines to learn in the vicinity of classification. Inspired by the functionality of biological neurons in human brain, the neural layers in CNN mimics the human brain. Neural layers in the CNN extracts the features of the data, image that we are providing as input. It outperforms the traditional machine learning techniques in feature extraction. The features learned by the network layers of CNN can be further used for the classification by using any classification algorithms. Each layer in the network learns different features from the input image. Deeper the layers more the features will be learned. Deep learning is the advancement of CNN in its number of layers and has improved the efficiency of the CNN. The deep network model, ALEXNET trained on ImageNet dataset

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

improved the efficiency and accuracy of the traditional image classification.

Transfer Learning in deep learning is an important and popular concept in the field of image classification and natural language processing. Here we choose a pretrained model and train it for a new problem. In deep models it requires many epochs or iterations to train the complete model and get higher accuracy. With pretrained models it takes only few iterations to train a model and get a desired accuracy. Thus, transfer learning saves a lot of computational power. The pretrained models are already trained with large dataset and we can use the weights of these models to train for a new problem even with lower dataset by modifying the last few layers of them. In one of paper of Manali et al they developed a model by fine tuning the VGG16 and VGG19 model with CalTech256 and GHIM10K database. Many works are being carried out on image classification using different pretrained deep models but with state -of-art databases. In our work we have developed a pretrained model with high accuracy to classify our own data set.

3 Proposed Work

In this section we describe the methodology implemented in classifying the different scenes and the overall steps involved in our work is shown in the block diagram. Majority of the deep learning models are trained on the publically available dataset. Here we have created our own dataset which we have collected from various websites. The dataset is a customized small dataset and we are using the concept of transfer learning approach to classify our scene dataset. In our work we have used the deep ResNet transfer learning model for feature extraction and classification. Here we evaluated performance of two variants of deep ResNet model, ResNet -50 and ResNet- 101 for our dataset.

4 Transfer Learning- ResNet

In deep neural network we always face a problem of vanishing gradient while updating the weights because we need to use back propagation while updating the weights and we use chain rule of calculus. The repeated multiplication will make the weights extremely small while they reach through the earlier layers. While updating the weights each layer is dependent on the previous layers. This problem of vanishing gradient is having a solution in residual network. ResNet standpoints for Residual Network. They have introduced a new concept known as skip connection in the network. Traditional network layers are learned with the output coming from each preceding layer while in residual network each layer is learned with the input applied also and we want to make input equal to output. Different variants of residual network we are discussing here are ResNet-50, ResNet -101.

ResNet-50: The value 50 means there are 50 layers in our network. The first layer is input layer having 7 x 7 filter and 64 such filters with stride 2. Next is a 3 x 3 max pooling layerwith stride 2. The first block of 1 x 1 filters with 64 such filters, 64 of 3 x 3 filers and 256, 1 x 1 filters. Second block with 4 layers each 128 numbers of 1 x 1 filters, 128 of 3 x 3 filters and 512 of 1 x 1 filters.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

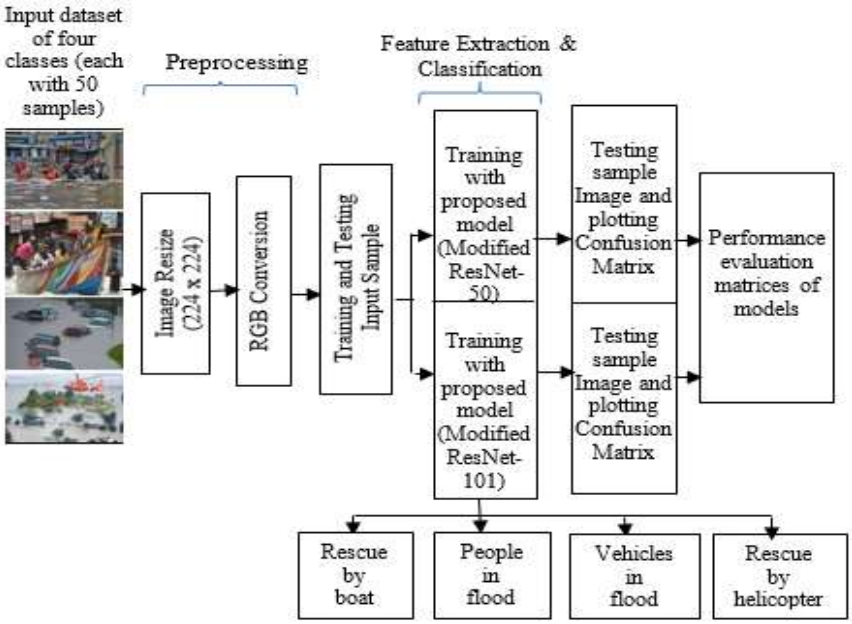

Fig. 1. Proposed model for scene classification using ResNet-50 and Resnet-101

Third block with 6 layers of 256 of 1 x 1 filters, 256 of 3 x 3 filters and 1024 1 x 1 filters. Forth block of 3 layers of 512 of 1 x 1 filters, 512 of 3 x 3 filters and 2048 of 1 x 1 filters. Next layer forms average pooling and fully connected layer. Altogether a total of 50 layers.

ResNet-101: This model architecture is having 101 layers by adding more 3 layer blocks in the network. The structure of ResNet -101 is similar to that of ResNet-50 having 24 repeated layers consisting of 256 of 1 x 1 filters, 256 of 3 x 3 filters and 1024 of 1 x 1 filters. It is here the difference lies in the architecture of Resnet-50 and ResNet-101. Remaining layers same as ResNet-101 thus forming a total of 101 layers.

5 Proposed Methodology

In our proposed methodology we have separated all the images into four separate classes. The images are then resized to 224 x 224 to be handled by the input layer of the ResNet models. These resized images are then preprocessed for RGB conversion. Preprocessed images are splitted into training and testing samples. These samples are then used to train the ResNet models. In our model the we have used ResNet-50 and Resnet-101 to extract the features from the samples. The deeper layer of the ResNet learns the sample dataset by learning many features and feature extraction is carried out by these deep models. Once the models have learned features we have to classify the images according to the learned features. For this we have made changes to the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

final few layers of ResNet-50 and ResNet-101. Since our classification is of four classes we have modified the final 3 layers of the ResNet. For our dataset we have replaced the last 3 layers with our own layers for carrying out the classification. The feature learned modified structured Resnet-50 and ResNet-101 are then evaluated using a test sample image. The confusion matrix is plotted for both the architectures and the performance parameters are evaluated.

5.1 Dataset

We have developed a dataset consisting of four classes of flood affected areas with 50 scenes in each class. The different types of scenes are rescuing of people in boat, people stuck in flood, Vehicles stuck in flood and rescuing by helicopter. We have collected these scenes from various sites and created our own dataset with an intension of assisting the people and the officials in finding the disaster affected areas and aid in rescue operations where there may be areas which are remote and non-reachable by people. Identifying these scenes by robots can enhance the efficiency of our application. We have trained this dataset by using the pre-trained deep ResNet architecture.

5.2 Pre-trained Models

In our work we have used pre-trained ResNet -50 and ResNet-101 for feature extraction. The input dataset is preprocessed to 224 x224 to be handled by the input layers of ResNet. The lower layers of ResNet is freezed and kept as the same architecture. These layers which are already pre-trained with the large image dataset can now easily learn our small dataset in less time with high accuracy. In both ResNet-50 and ResNet- 101 we have customized the last three layers with our layers: Fully connected layer of 4, Soft Max Layer and Classification Layer for 4 classes

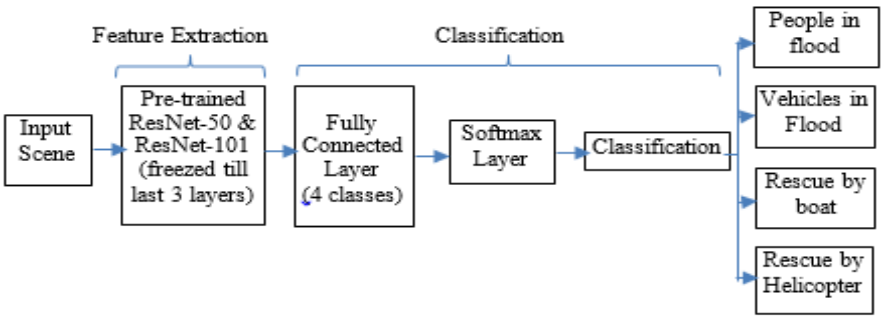

Fig. 2. Proposed architecture of ResNet-50 and ResNet-101 for scene classification

6 Experimental Results

We have done the experiment with a small dataset of 200 scenes with 50 scenes separated into four classes. Each input scene was resized to 224 x 224 and carried out RGB conversion to get suited with the input layer of ResNet architecture. 70 % of

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

preprocessed dataset divided into training set samples and remaining 30% to test set samples and trained by the pre-trained ResNet 50 and ResNet 101 models whose last layers were modified for our application. We have trained the network with batch size of 15 and for 20 epochs and each model is tested with sample input. The execution time for both models are evaluated. We have measured the efficiency of both the models by plotting theconfusion matrix. The different performance matrices like accuracy, Precision, F1- Score also evaluated for both architectures.

6.1 Training Progress - ResNet50 & ResNet 101

The input samples were trained by the ResNet for a batch size of 15 and 20 epochs. Training progress monitored and the accuracy and lose model graph is plotted. After successful training the model is tested with test sample with ResNet50 having an elapsed time of 52.488234 seconds and ResNet 101 of 106.29 seconds.

6.2 Performance Metrics

We have evaluated the two proposed models with their performance matrices. We obtained the Tp (True Positive) value indicating the correct prediction of true class, Tn (True Negative) value indicates the correct prediction of false class, Fp (False Positive) and Fn (False Negative) both showing the wrong prediction of classes.

Accuracy = True positive + True negative / (Tp +Tn + Fp+ Fn) *100 (1)
Sensitivity = True positive / (True positive + False negative) *100 (2)
Specificity = True negative / (True negative + False positive) *100 (3)
Precision = True positive / (True positive + False positive) (4)
Recall = True positive / (True positive + False negative) (5)
F1-Score = (2*Precision*Recall) / (Precision+Recall)*100 (6)

We have used these parameters to evaluate the performance parameters of the two models and we have compared the two models.

Table 1. **Performance evaluation of proposed ResNet50 and ResNet101**

Performance Matrices	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
Proposed Model (ResNet50)	89.28	85.71	92.85	0.92	0.85	88.88
Proposed Model (ResNet101)	91.429	87.143	95. 71	0.95	0.87	91.04

7 Conclusion

We have proposed our model with an objective to classify natural scenes where we have combined different scenes. We collected 200 scenes and grouped them into four different classes indicating rescue byboat, people in flood, vehicles in flood and rescue

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6

by helicopter. We have proposed such a model to assist the robots and drones to identify the remote location affectedby flood which we people may not notice. Both ResNet50 and ResNet101. models areevaluated for their performance matrices- Sensitivity, Specificity, F1-Score etc. ResNet50 showed F1-score of 88.88 and ResNet101 showed 91.04. The execution time and accuracy of both models were also compared. Resnet50 gave an accuracy of 89.3% with an execution time of 52.4 sec and Resnet101 showed the accuracy of 91.4% with a slightly larger execution time of 106.29 sec. Both the proposed models gave an appreciable good response even with the small dataset considering execution time and accuracy. As a future scope we develop our model with real time scene images.

References

[1] Kamavisdar, P., Saluja, S., & Agrawal, S.: A survey on image classification approaches and techniques. International Journal of Advanced Research in Computer and Communication Engineering, 2(1), pp. 1005–1009 (2013).
[2] Sharma, M., Anuradha, J., Manne, H. K., & Kashyap, G. S. C.: Facial detection using deep learning. IOP Conference Series: Materials Science and Engineering, 263, 042092 (2017).
[3] Ardhendu Bandhu, Sanjiban Sekhar Roy.: Classifying multi-category images using Deep Learning: A Convolutional Neural Network Model, 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology, pp. 1-6 (2017).
[4] S.Regina Lourdhu Suganthi, Hanumanthappa, S. Kavitha.: Event Image Classification using Deep Learning," in 2018 International Conference on Soft-computing and Network Security, pp. 11-17 (2018).
[5] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton.: Deep learning, Nature 521.7553, pp. 436-444 (2015).
[6] Schmidhuber, Jürgen.: Deep learning in neural networks: An overview, Neural networks, pp. 85-117 (2015).
[7] Weng, Qian, et al.: Land-Use Classification via Extreme Learning Classifier Based on Deep Convolutional Features, IEEE Geoscience and Remote Sensing Letters (2017)
[8] Chauhan, K., & Ram, S.: International Journal of Advance Engineering and Research Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensorflow and Keras, pp. 533–538 (2018).
[9] Zhang, F., Du, B., & Zhang, L.: Scene classification via a gradient boosting random convolutional network framework. IEEE Transactions on Geoscience and Remote Sensing (2016).
[10] Santisudha Panigrahi, Anuja Nanda, Tripti Swarnkar.: Deep Learning approach for Image Classification, in 2018 2nd International Conference on Data Science and Business Analytics (2018).
[11] Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, IEEE Transactions on Geoscience and Remote Sensing, 54(10), pp. 6232–6251 (2016).
[12] Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., & Emery, W. J. : SVM active learning approach for image classification using spatial information. IEEE Transactions on Geoscience and Remote Sensing, 52(4), pp. 2217–2223 (2014).
[13] Korytkowski, M., Rutkowski, L., & Scherer, R.: Fast image classification by boosting fuzzy classifiers. Information Sciences, 327, pp. 175–182 (2016).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_6