Special Issue - 2022

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTCSIT - 2022 Conference Proceedings

Forest Fire Detection using Combined Architecture of Separable Convolution and Image Processing

Navyashree J[1], Veena M Naik[2]

[1,]Student ISE Department, Sri Krishna Institute of Technology, B'lore-560090, India

[2] Faculty ISE Department, Sri Krishna Institute of Technology, B'lore-560090, India

Abstract: Due to the recent record-setting wildfire outbreaks throughout the world, early identification and categorization of wildfires utilising aerial image-based computer vision algorithms such convolution neural networks and image processing approaches has gotten a lot of interest. Previous research has shown different degrees of success in developing forest fire classification algorithms utilising variations of well-known complex convolutional neural network designs, which demand a lot of training time but have low predictive power and high false alarm rates. This research proposes a hybrid architecture of separable convolution neural networks and digital image processing employing thresholding and segmentation to accurately identify small-scale forest burning, which generally signal the start of more catastrophic catastrophes. The suggested design is straightforward, making it less computationally costly. Evaluation of performance based on test results.Object Detection, Navigation, Wireless Charging, Water Sensor.

Index terms:- Forest fire, image processing, deep learning,seperable CNN, image classification, thersholding.

I. INTRODUCTION

In the United States, record-breaking conflagrations in California, Washington, and Oregon threatened public life, the local economy, and electrical infrastructure in 2020. Given the variability in rainfall patterns, climate- induced droughts, and human activities in sensitive wildlife-urban fringe zones, the problems are projected to worsen [1], [2]. As new grid reliability indicators and innovative grid hardening approaches are deployed, electrical utilities with aged infrastructures operating in critical fire zones are increasingly subjected to heightened regulatory scrutiny. In tandem with early- warning applications, research interest has switched to resilience studies focused on high-impact, low- probability occurrences [3], such as grid-induced wildfires and grid operations during severe weather.

Firefighters and first responders can assist control wildfires by predicting fire behaviour and discovering them early on. To aid local and regional fire departments in planning for the season, the authors of [4] advocated constructing seasonal statistical models and integrating them with regional risk assessments. The authors of [5] suggested visualising an optimum power shut-off issue, which would allow grid operators to optimise transmitted power while selectively de-energizing.components of the electric grid to reduce the danger of wildfire ignition [6] have observed some success with the use of appropriate placements of specialised cameras in conjunction with an intelligent video smoke detection system.

Artificial intelligence-based CNN appears to be at the vanguard of the numerous new technologies available to identify wildfires in their early phases. Data from wildfires acquired by UAVs and satellites using optical and thermal imaging systems can help with precision neural network training [7]–[11]. Advanced CNN architectures with complicated problem-solving capabilities, such as AlexNet, GoogLeNet, LeNet, ResNet, and U-Net, emerged as forerunners as computer vision progressed [12], [13]. However, when applied to forest fire categorization, the recommended designs are computationally costly, with limited specificity and large false alarm rates in the presence of smoke. For example, in [7], the authors examined many unique ways to solving the fire picture classification problem, and modified GoogLeNet produced the best results, with 96.9% training accuracy and a 96.9% recall rate.1.5 hours of training time for just 289 photos In [8, the authors used a modified U-Net architecture to achieve 91.99 percent accuracy, 83.88 percent recall, 87.75 percent F-1 score, and 99.85 percent AUC in the training set for fire segmentation. When trained with 39375 test-bed photos, GoogleNet and U-Net proved computationally costly, taking more than fifteen epochs and around 1500 seconds each epoch to achieve an accuracy of better than 90%. Other designs, such as Alexnet, LeNet, and ResNet, used more than 1000 seconds each epoch and had a validation accuracy of less than 79 percent after fifteen epochs.

A combination architecture comprising a basic separable CNN model with regularisation and a digital image processing unit with thresholding and segmentation is proposed in this work. Furthermore, the basic separable CNN model is substantially less computationally expensive (training time of 1.36 hours for 39375 photos) and has an AUC of 100% on the training set and a 97 percent AUC with a 92 percent F-1 score on the test set. Overall, the suggested integrated architecture was able to reliably predict forest fires with a sensitivity of 98.10 percent and a specificity of 87.09 percent, outperforming complicated algorithms.

The remainder of the paper is arranged as follows: Section III provides a detailed description of the study's dataset as well as background information. The methodologies used in this study for forest fire categorization are highlighted in Section IV.

The simulation results and performance evaluation measures are shown in Section V. Finally, Section VI discusses closing remarks and future ramifications.

II. STUDY BACKGROUND AND DATA DESCRIPTION

The dataset for this investigation came from IEEE DataPort, which includes UAV-captured aerial images from a managed pile fire in the Northern Arizonian pine forest in the United States [14]. The static repositories were prepared by the participating authors.by preparing the in-situ video data, comprising training/validation dataset and test dataset pictures The dataset photographs, which are typical of the geographical location [15], [16], include dense undergrowth, scrubs, lakes, riverbank, sunsets, and sunsets with varying levels of snow in the background.

In predictive modelling, geographic characteristics are critical [17]. The research area's topography attributes assist narrow down the areas where the defined model may be successfully implemented. The US states of Texas, California, Oregon, and Arizona, for example, are among the high-risk forest fire zones, as shown in [4]. CNN architectures employing aerial images from Northern Arizona can be deployed in the US states of California and Oregon due to geographic similarities. Due to Texas's unique geography, the same neural network may require additional training for judging forest fires.

The initial simulation and modelling requirements for this study were met with Python 3.8.5 on a typical Windows PC with an Intel i5 CPU and 12 GB of RAM. With Python's Scikit-Learn based GridsearchCV, Google Colab Pro was utilised to allow full-scale model performance tuning. Tensorflow 2.0, Keras, OpenCV 3.4.13, and Scikit-Image 0.18.1 were among the Python packages utilised in the simulation research.

III. METHODOLOGY

A. Neural Networks, CNNs, and Separable CNNs Overview

The CNN algorithm tries to mimic the visual cortex of the eye, which divides pictures into discrete receptive fields that respond to particular spots in the visual field [12]. Because of its unique spatial independent technique, which does not need using all available pixels from an input picture in the early layers, CNN takes less time to compute than deep neural networks [18]. A basic convolution operation can be stated mathematically as in CNN employs a subsampling layer known as the pooling layer, which collects or takes the greatest value from the input data to prevent overfitting. The input layer of a CNN architecture is followed by the output layer A CNN version is separable convolution neural networks.

A separable CNN architecture, unlike a traditional CNN design, represents spatial and cross-channel characteristics (such as intricate patterns) independently for each available colour channel. The depthwise convolutional process is used to extract spatial features in a separable CNN, and it can be stated mathematically as in while the extraction of cross-channel features is performed using pointwise convolution operation as shown in (2) and (3), image characteristics are examined using two separate techniques at the same time. Separable CNN is computationally quicker and has superior classification properties than previous CNN designs [12]. [19], [20] provide an excellent exposition of the working theory of depthwise and pointwise covolution.

B. Thresholding and Segmentation Preview

Segmentation, as the name implies, seeks to separate or segment comparable items in a raw picture based on texture, colour, and brightness uniformities [21]–[23]. One of the most common ways of segmentation is thresholding. Each pixel in an image is compared to the supplied threshold value T in a basic thresholding approach, also known as binary thresholding. Pixel values (pixelvalue) that are less than or equal to T are set to zero, while those that are equal to or more than T are set to a user-specified maximum value. The same threshold value T is applied to each channel one at a time in multichannel (or coloured) pictures. The result is a single merged image made up of thresholded images from several channels. Another simple yet effective object extraction approach is to mention a range of HSV values. Using the HSV colorspace thresholding approach [22], [24], it is simple to extract objects with similar colour and brightness.

HSV values set by the user serve as lower (Tmin) and upper (Tmax) threshold values. Outside of Tmin and Tmax, pixel values (pixelvalues) are set to 0. Object identification in grayscale photos using HSV colorspace may be stated mathematically as pixelvalue = pixelvalue

, Tmin ≤ pixelvalue ≤ Tmax.

C. Techniques Adopted

Computationally costly implementations of well-known CNN architectures like AlexNet, Lenet, and ResNet (greater than 1000 seconds per epoch). Even though the training accuracy of each of these algorithms was better than 99 percent, the test set (or validation) accuracy barely exceeded 79 percent. Overfitting was discovered as the root cause of such inaccuracies. Further examination of the misclassified photos revealed the presence of smoke and fog in both the fire and no fire. Classifications.Misclassification looked acceptable because smoke or fog may largely alter the hue and saturation of a raw picture. The remainder of this section provides a thorough summary of the strategies used to combat overfitting and smoke/fog.

1) Countermeasures against Overfitting: When compared to complicated models with numerous convolution layers, a simple CNN model with two separable two-dimensional convolution layers performed well. When compared to a normal two-dimensional convolution network, a model with separable convolutional layers demonstrated improved accuracy with less computing time. Performance parameters such as AUC, accuracy, precision, recall or sensitivity, and specificity versus the test set were used to finalise the CNN model. The performance of models with one, two, three, four, and six convolution layers was also evaluated in terms of validation accuracy and computational time. The max-pooling layer, followed by each of the two

Special Issue - 2022

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTCSIT - 2022 Conference Proceedings

convolution layers, produced better results than the average pooling method. The accepted model performed best with a kernel size of utilising the ReLU activation function of 3X3.

In addition to choosing a less complicated neural network, the model included three types of regularisation techniques to significantly reduce overfitting: L2 regularisation in each convolution, batch normalisation, and a dropout layer before the output layer, as well as data augmentation and early stopping. Each of these regularisation methods is described in detail below.

• Ridge regression is used in the L2 regularisation, which reduces the total MSE by reducing variance and introduces a tiny bias.

• To avoid vanishing/exploding gradient concerns, batch normalisation is frequently included in a model.The batch normalisation approach, as the name implies, normalises and zero-centers each input (6c), then scales and shifts the result to aid the algorithm in learning the best mean and scale for each input layer (6d)The batch normalisation algorithm is described in [12].Though model complexity and slower predictions are

often associated with the iterative nature of batch normalization, however, the learning is eventually sped up[25] which inturn imposes a regularizing effect that reduces overfitting. A more detailed overview of batch normalization is available in [12].

• Finally, a drop out layer is added just before the output layer. The dropout layer facilitates better generalization. The hyperparameter (p) in a dropout layer suggests the probability of a neuron being dropped at a layer is p for every step during training [12].

• Data augmentation is yet another technique to prevent overfitting and improve model performance in imbal [26]–[28]. anced classes Data augmentation increases the size of the training set artificially by generating randomly changed data from existing data, which helps the learning process by supplying additional trainingdata.
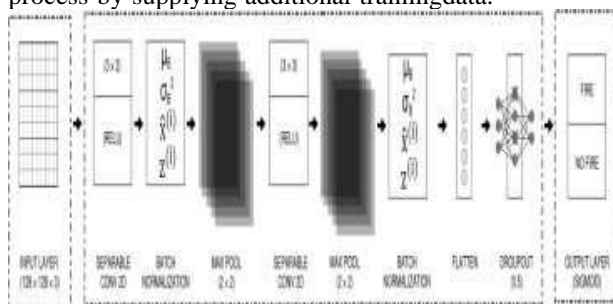

Fig. 1: Architecture of the proposed simple separable convolution neural network

• By analysing the validation error or validation accuracy, early halting regularises iterative learning. As soon as the validation error or validation accuracy begins to rise, the learning process from the training set is terminated.

The detailed construction of the proposed separable CNN architecture combined with measures to mitigate overfittingis shown in Fig. 1.

a)   *Raw digital image files*

b)   *Applying multichannel binary thresholding*

2) Smoke and Fog Measures: Thresholding and segmentation were successful in combating smoke and fog misclassification. Multichannel binary thresholding is followed by segmentation utilising two sets of HSV colorspace filters in the digital image processing approach suggested in this research. The use of thresholding and segmentation to recover a desired portion of the raw picture is shown in Figure 2. Two raw photos, one with two pile burns or brushfire and the other without a neighbouring pile burn, are shown in Fig. 2(a). Multichannel binary thresholding output is shown in Figure 2(b). The minimum value is 100, while the maximum value is 1000.is set to 255. The ultimate result of the digital image processing is shown in Figure 2(c). The pile burn area is extracted using two sets of HSV colorspace filters. The difference between fire and no fire is determined by whether any pixel values other than 0 (0 specifies the colour black) were found in the final output image, as in the case of fire.

Despite the fact that the basic separable CNN model performed well in recognising fire incidents, smoke and fog remained an issue. Image processing might effectively manage the smoke and fog problem. As a result, as illustrated in Fig. 3, a combined architecture with basic separable convolution and image processing was developed. The suggested architecture takes the results of both models and assigns a final verdict: (1) fire if both models predict fire; (2) no fire if both models predict no fire; and (3) unsure if just one of the two models predicts fire. Human assistance was necessary for raw photos that were uncertain. Special attention was paid to the performance such that the undecided category did not surpass 2% of the total.

IV. RESULTS OF SIMULATION AND DISCUSSION

This section provides a comprehensive analysis of the proposed model metrics and assesses model performance utilising Keras TensorFlow optimizers such as Nadam, Adam, Adagrad, SGD, Ftrl, and RMSprop. In addition, a fast demonstration of how to tune hyperparameters such batch size and learning rate is provided.
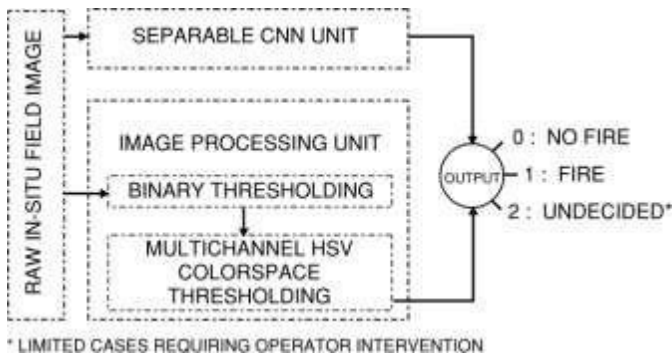
Fig. 3: *Proposed combined architecture using separable CNN and digital image processing*

A. CNN Performance Assessment

Section IV-C discusses the steps needed to arrive at a simple CNN model with acceptable performance. However, the proposed CNN model was chosen from among numerous variations by optimising hyperparameters such Using Scikit-GridSearchCV, learn's you may control batch size, learning rate, and Keras optimizer. GridSearchCV tested the following sets of hyperparameters, with the highest performing values from each set indicated in boldface: Keras optimizer: [Nadam, Adam, Adagrad, SGD, Ftrl, RMSProp], Batch size: [32, 64, 128], Learning rate: [0.01, 0.001, 3e-4], Keras optimizer: [Nadam, Adam, Adagrad, SGD, Ftr Table I shows that the choice of an optimizer had the most impact on model performance among the hyperparameters given. With a learning rate of 0.001 and a batch size of 64, Adam optimizer provided the greatest overall performance among the various combinations. On Google Colab Pro, training the chosen basic CNN model with 39375 photos took eleven epochs with an average computation time of 450 seconds per epoch. In comparison, the model's performance was examined on a normal Windows i5 PC, which took 151 seconds to compute while examining 8608 photos. The overall accuracy of the simple separable CNN model was 92 percent, with a sensitivity of 91.26 percent and a specificity of 92.51 percent

VI. CONCLUSION

The proposed combination of simple separable CNN and image processing for fire picture classification produced an effective and accurate outcome. The model outperforms the sophisticated predecessors with a false alarm rate of less than 1.90 percent and a specificity of 87.09 percent. Because the number of photos marked as unsure remained below 2%, human interaction can be used to improve the overall model performance.

In the event of a pile fire, trained architectures such as the one proposed in this research can be pre-loaded into the drone memory, triggering alarms on the first responder's system. Furthermore, the drone system's GPS coordinates can be used to verify immediate actions..

Because the acquisition of location-specific photos is a prerequisite for training a CNN model, significant study attention should be paid to how such images might be obtained artificially or without causing safety problems.

Combining a certain degree of topographical elements that overlap between locations is one such technique.

REFERENCES

[1]. A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z.Fule, and E. Blasch, "The flame dataset: Aerial imagery pile 'burn detection using drones (UAVs)," 2020. [Online].

[2] A. E. Maxwell, P. Pourmohammadi, and J. D.Poyner, "Mapping the topographic features of mining-related valley fills using mask R-CNN deep learning and digital elevation data," RemoteSensing, vol. 12, no. 3, p. 547, 2020.

[3] L. Dang, P. Pang, and J. Lee, "Depth-wise separable convolution neural network with residual connection for hyperspectral image classification,"Remote Sensing, vol. 12, no. 20, p. 3408, 2020.

[4] J. W. Muhs, M. Parvania, and M. Shahidehpour, "Wildfire risk mitiga-tion: A paradigm shift in power systems planning and operation," IEEE Open Access Journal of Power and Energy, vol. 7, pp. 366–375, 2020.

[5] T. Wang, A. Li, W. Xu, J. Yang, and Z. Zhang, "The applied research onWUI fire risk prevention and control," in 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication(ICEIEC). IEEE, 2020,pp. 285–288.

[6] S. Ghosh and S. Dutta, "A comprehensive forecasting, risk modelling and optimization framework for electric grid hardening and wildfire prevention in the US," International Journal of Energy Engineering,vol. 10, no. 3, pp. 80–89, 2020