International Journal of Signal and Imaging Systems Engineering

ISSN online: 1748-0701 - ISSN print: 1748-0698
https://www.inderscience.com/ijsise

Fire detection in nano-satellite imagery using Mask R-CNN

Aditi Jahagirdar, Neha Sathe, Sneh Thorat, Saloni Saxena

DOI: 10.1504/IJSISE.2024.10060648

Article History:

Received:	15 February 2023
Last revised:	15 April 2023
Accepted:	09 October 2023
Published online:	15 July 2024

Fire detection in nano-satellite imagery using Mask R-CNN

Aditi Jahagirdar*, Neha Sathe, Sneh Thorat and Saloni Saxena

School of Computer Engineering and Technology,
Dr. Vishwanath Karad MIT World Peace University,
Pune, 411038, India
Email: aditi.jahagirdar@mitwpu.edu.in
Email: neha.sathe@mitwpu.edu.in
Email: 1032190137@mitwpu.edu.in
Email: salonisaxena556@gmail.com
*Corresponding author

Abstract: Increasing availability of satellite imagery has made it possible to detect forest fires through satellite imagery. This research aims at investigating early forest detection approaches using deep learning and satellite image segmentation. The algorithms implemented in this work are mask region-based convolutional neural network (Mask R-CNN), UNet and deep residual U-NET (ResUNet). The experimentation is carried out on publically available satellite image data having challenges like the presence of clouds, snow, rivers and sand, which gets confused with the smoke from the fire. The methods implemented here can successfully distinguish between these natural entities and the smoke emitted from the fire. It is seen that Mask R-CNN has an IoU of 0.925, whereas UNet and Res-UNet have IoUs of 0.30 and 0.35, respectively. The results indicate that Mask RCNN is both more time effective and precise and can be used in forest fire detection systems.

Keywords: satellite images; image segmentation; Mask R-CNN; ResUNet; UNet; deep learning.

Reference to this paper should be made as follows: Jahagirdar, A., Sathe, N., Thorat, S. and Saxena, S. (2024) 'Fire detection in nano-satellite imagery using Mask R-CNN', *Int. J. Signal and Imaging Systems Engineering*, Vol. 13, No. 1, pp.19–26.

Biographical notes: Aditi Jahagirdar received her Bachelor's degree in 1993 and a Master's degree in 1999 from the University of Pune. She completed her PhD in the Domain of Computer Vision in 2021 from Savitribai Phule Pune University, India. She is currently working as an Associate Professor in the School of Computer Engineering and Technology, at MIT World Peace University, Pune. She is having a total of 27 years of teaching experience In addition to publishing more than 40 papers, she has two patents awarded to her. She is a Member of the ACM, IET, CSI and IETE. Her research interests include computer vision, Image processing, video processing, and pattern recognition.

Neha Sathe received her Master's in Engineering in 2012 from Maharashtra Institute of Technology, Pune affiliated to Pune University in Digital Systems. She completed her PhD from MIT World Peace University, Pune, India, and working as an assistant professor in the School of Computer Engineering and Technology. Her areas of interest are image and signal processing, wireless sensor networks, and data analysis.

Sneh Thorat is pursuing a master's in Computer Science at USC after completing a Bachelor's at MIT World Peace University. His research areas encompass computer vision, natural language processing, and machine learning, with demonstrated proficiency in object detection, image segmentation, syntactic, and semantic analysis. He possesses a fervent passion for deciphering the intricate mathematics at the core of machine learning and aspires to make substantial contributions to the field by developing new algorithms that can address real-world challenges and benefit society.

Saloni Saxena is a Computer Science Engineering student at MIT World Peace University who delves into research areas such as evolutionary algorithms, generative AI, and computer vision while looking forward to expanding her knowledge of artificial intelligence and the intricacies of data. Her experience extends to research in object detection and segmentation, genetic algorithms, and AI solutions for various facets of computer science, such as systems and computation. She enjoys unravelling the complex mathematics at the core of machine learning and looks forward to making meaningful contributions to the field.

1 Introduction

Wildfires are the most pervasive hazard in forests. Forest fire is an unavoidable catastrophe that directly or indirectly affects flora and fauna of not only the area it covers but beyond that as well (Robinne, 2021). The disastrous bushfires in Australia, for example, have caused approximately 800 deaths in people and billions in animals. The after-effects of such fires lead to more deaths due to habitat loss. Such fires can also cause severe air pollution at a local level. Thus, early detection of fire hazards is vital to avert a large-scale disaster.

Forest fires were essentially detected by human observations from lookout towers (Panagiotis et al., 2020). Unsurprisingly, this fire detection method had many areas that demanded reform, which sparked further development. Fire can be detected with sensors. However, they are not very efficient for the early detection of fires as particles take time to reach the sensors and activate them (Krüll et al., 2012). As a conventional means of detection, however, it is evident that even sensors and their other contemporaries had imperfections. Consequently, an entirely new approach to fire detection became imminent. With advancements in computer vision, machine learning, and remote sensing, new tools were developed that overcame most of the hindrances of previously used methods. Satellite images can be used as a reliable dataset for such devices. The use of satellite images enables more frequent or constant monitoring of large areas of forests. Nowadays, wildfires are occurring more frequently and severely as a result of climate change. The need for improved tools to monitor and lessen the effects of wildfires has spurred this research work.

The task of detecting fires from aerial imaging data is difficult and laborious. The ability to identify smoke in an image is mostly what allows for fire detection. The satellite images of the forest area have a fire, clouds present along with the smoke. The structure of the cloud and smoke is very identical which makes it challenging to separate them. For better detection of fire, the approach used must be able to differentiate between cloud and smoke. Additionally, the background, i.e., parts of the image other than the smoke, change to a great extent over different terrain.

In this study, semantic segmentation based on deep learning is applied to nano-satellite images to detect fires. Here, Mask R-CNN, a method that has produced good results in other applications, is implemented. The technique creates a pixel-wise mask of each image object. By using this technique, it is possible to gain a deeper understanding of the objects in an image. The results are compared with the results obtained by implementing UNet and ResUnet. The results show that, Mask R-CNN can successfully distinguish smoke from clouds.

2 Literature review

The literature shows various methods used by researchers for fire detection. The three major steps used in the conventional approach for fire detection are fire detection, feature extraction, and image pre-processing. Recently various deep-learning approaches have been implemented for fire detection. This paper proposes using Mask RCNN (He et al., 2020) and ResUNet (Foivos et al., 2020) algorithms for fire and smoke segmentation. Mask RCNN is a deep neural network consisting of predicting the object's position in an image and concentrating the bounding box and producing a mask of the object at a pixel level. ResUNet is the next and better version of UNet (Zhang et al., 2018). UNet is a convolutional network that includes down-sampling and up-sampling of features. After implementing these algorithms, we have compared the results and accordingly concluded which algorithm is more reliable and effective for fire detection.

Various approaches have been suggested by researchers for tackling these problems. The segmentation technique (Kumarguru and Siau-Chuin, 2015) includes the detection of the colour and edges of the fire. For detecting the edges, the Sobel edge detection algorithm is implemented. The colour and edges together are used to identify the region of interest. The task of determining smoke from the same patterns such as seaside, dust, haze, etc., is demonstrated in Rui et al. (2019). An advanced convolution neural network model – SmokeNet is proposed, which embodies channel-wise and spatial in the Convolution neural network. An approach for fast and accurate recognition of smoke from a video is presented in Gagliardi et al. (2021). Both AI and Image processing methods are encompassed in this study. The proposed CNN was fast and lightweight. The F1 score was 0.8837, and the accuracy was 0.8438. An architecture that can work on very few training images and still provide precise segmentation is proposed in this paper (Ronneberger et al., 2015). A large number of feature channels are used in this u-shaped architecture. To apply this network to large images, an overlap-tile strategy is used. The network gives good results in different biomedical segmentation applications and can be used beyond that, too. A framework, for instance, segmentation extending Faster R-CNN by adding a branch that outputs the object mask was presented in Foivos et al. (2020). Mask R-CNN improves upon Faster R-CNN by making it simple to implement and train. The mask target is the junction between a Region of Interest and its ground- truth mask. Comparing Mask R-CNN and the state-of-the-art methods reveals that Mask R-CNN outperforms the other models and gives good results even under challenging conditions.

Li and Zhao (2020) Proposes novel image fire detection algorithms based on advanced object detection CNN models. The algorithms achieve higher accuracy and real-time detection capabilities compared to traditional detection methods. Among the four proposed algorithms, the YOLO v3-based algorithm achieves the highest accuracy of 83.7%, and is also the most robust and fastest, with a detection speed of 28 FPS. The study concludes that using object detection CNNs for fire detection in images is feasible and effective.

In Saeed et al. (2020), a multimodal strategy based on CNN is proposed. Using data from sensors and images, it employs machine learning and deep learning algorithms. Adaboost- MLP for fire prediction, Adaboost-LBP for emergency detection and ROI generation, and convolutional neural network (CNN) for fire detection utilising pictures from security cameras are the three deep neural networks that are suggested in this paper. The proposed model has a low rate of false alarms and 99% accuracy. The research concludes that the suggested strategy is more effective than conventional ways and can be strengthened even more with additional training.

Huang et al. (2022) proposes a novel fire detection method, called Wavelet-CNN, that combines wavelet transform and CNNs to improve accuracy and reduce false alarms in video surveillance. The method extracts spectral features of the image using the 2D Haar transform and inputs them into CNNs at different layer stages. ResNet50 and MobileNet v2 (MV2) are used to test the method on benchmark datasets, and results show that it improves detection accuracy and reduces false alarms, particularly for the lightweight MV2. The method achieves comparable accuracy to state-of-the-art methods with low computational needs.

A survey of various methods is given in Li et al. (2021). This review paper provides a comprehensive survey of CNNs, including their history, various convolutions, classic and advanced models, related functions, applications, and prospects. The paper also presents guidelines for devising novel networks, including accuracy and speed considerations, and discusses activation functions, loss functions, optimisers, and hyperparameter selection. Applications of 1D, 2D, and multidimensional CNNs are covered, and open issues and promising directions for CNNs are discussed. The authors suggest that CNNs can be refined further in terms of model size, security, and NAS and discuss some hardware implementation schemes for CNNs. The study of Faster RCNN and Mask RCNN algorithms was carried out by studying several research papers that have used these algorithms for different applications.

A Faster RCNN and Mask RCNN-based approach for road crack detection (Xu et al., 2022) has been presented in this paper. The study showed that joint training strategies prove to cause problems for their particular use case. Results from Faster RCNN proved to be slightly superior to that of Mask RCNN for the same learning for most of the datasets they were tested on, however, at learning rates more than 0.005 Mask RCNN shows better accuracy than the former. Their method also works better than YOLOv3.

The study (Kim and Lee, 2019) proposes a Faster RCNN-based approach to detect suspected fire regions in long-term videos. This deep learning-based approach mirrors the manual fire detection method called DTA. The spatial features calculated by Faster RCNN are then passed on to a Long Short Term Memory Network which helps in the labelling process. This ensures that the temporal behaviour of the fire is checked and no false alarms are raised. The LSTM model gives an accuracy of 97.2%. However, it is shown that Faster RCNN often confuses true smoke with steam.

The researchers of Pan et al. (2020) suggest the use of Additive Neural Networks for fire detection from video data. Their results show a time-saving of 12.4% in comparison to traditional CNN-based detection. Thus their work aims to implement a deep neural network-based detection using a simple processor. Their model outperforms binary-weighted CNNs and yields and accuracy are only slightly lower than regular CNN. They propose that the solution they provide is more robust and thus more practically feasible.

It is evident from the literature review that image segmentation is essential for fire and smoke detection. In order to achieve better results, existing image segmentation techniques can be combined in an efficient manner. Techniques such as RCNN and Mask RCNN have proven to be effective in segmentation and can give good results in forest fire detection.

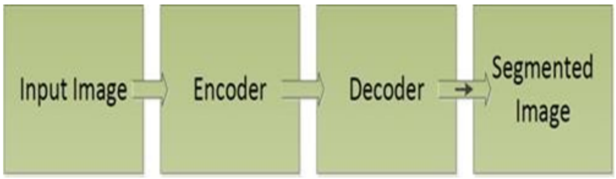
3 Methodology

In this work, mask region-based convolutional neural network (Mask R-CNN), UNet and deep residual U-NET (ResUNet) algorithms are implemented for detecting the forest fire from satellite images. All these methods use semantic segmentation based on deep learning.

Semantic Segmentation groups similar parts of the image belonging to the same class. Semantic Segmentation follows three steps: Classifying, Localising, and drawing a bounding box around it. For each item in the picture, image segmentation generates a pixel-wise mask. This method allows one to get a far more comprehensive understanding of the object(s) in the image.

Semantic image segmentation is a stage between crude and accurate inference. The origin might be found in classification, which entails creating a forecast for the entire input. The following stage is detection/localisation, which renders the classes and additional information about their geographical position. As a result, we frequently find popular image classification networks acting as a backbone, i.e., an encoder. They are commonly used as the foundation of semantic segmentation systems. Hence, an encoder network tailgated by a decoder network may be considered a generic semantic segmentation architecture. Figure 1 shows a block schematic for the semantic segmentation approach.

Figure 1 Block diagram of semantic segmentation (see online version for colours)



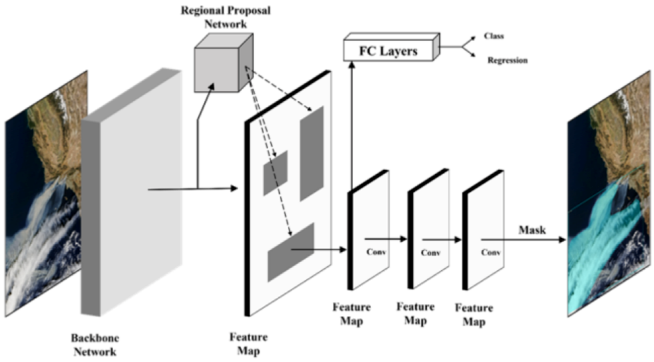
An encoder does convolution and down-sampling of the image while a decoder applies convolution and up-samples the image. A pre-trained classification network, such as VGG/ResNet, is generally used as the encoder, followed by a decoder network. The decoder's job is to effectively protect the encoder's learned discriminative features (lower resolution) onto the image pixels (higher resolution) to provide a dense classification.

3.1 Mask R-CNN

R-CNN (Regions with CNN feature) is an example of region-based Semantic Segmentation techniques. Based on the object detection findings, it conducts semantic segmentation. R-CNN uses selective search to extract many object suggestions before computing CNN features for each of them. Finally, it uses class-specific linear SVMs to classify each area. Unlike standard CNN structures, which are fundamentally designed for image classification, R-CNN can handle more intricate tasks like object identification and image segmentation. It may even form the grounds for both disciplines. The Faster R-CNN is an advanced form of R-CNN. Faster R-CNN is a type of region-based CNN that provides bounding boxes and a confidence score for each object's class label. The Faster R-CNN delivers a better job of incorporating an attention mechanism.

Mask RCNN is a Faster RCNN-enhanced instance segmentation system. It has two stages. The first stage *is* made up of two networks namely the backbone network (ResNet, VGG, Inception, etc.) and the region proposal network. These networks execute once for each image to provide a list of region predictions. The feature map's regions that contain the object are known as region proposals. For each predicted region that was obtained in Stage 1, the network predicts bounding boxes and item classes in Stage 2. While fully connected layers in networks always require a constant size vector to generate predictions, each proposed region can be of varied size. These suggested regions' sizes are set using the RoIAlign technique or RoI pool, which is quite similar to MaxPooling. Thus a mask is generated for each predicted region. This mask is a binary mask outputted for each region of interest. Figure 2 shows the architecture of Mask RCNN.

Figure 2 Mask RCNN architecture (see online version for colours)



Mask RCNN's multi-task loss function integrates classification, localisation, and segmentation mask losses, equation (1) where L_{cls} and L_{box} are like it was in Faster R-CNN

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

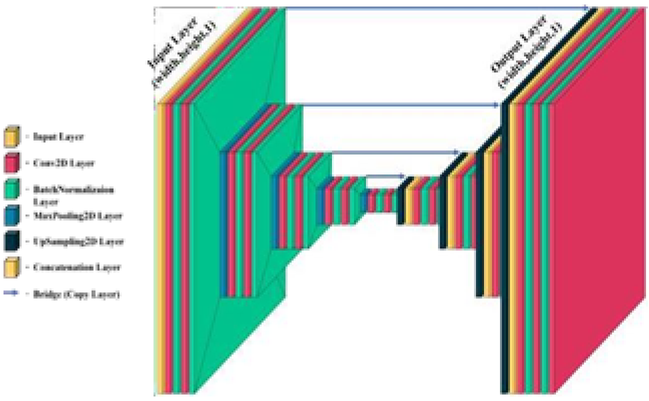
A mask of dimension $m \times m$ is generated by the mask branch for each class and region of interest (RoI). Since the model aims to learn a mask for each class, classes have no competition for generating masks.

3.2 UNet

UNet is an encoder-decoder architecture that utilises the power of FCN. UNet is not a pure encoder-decoder network, although it closely resembles an encoder-decoder design. The encoder and decoder have a clear distinction: the encoder converts each sample's representation into a 'code' in the latent space. The decoder may only generate outputs given such codes. This means that you can disassemble such a network and utilise the encoder and decoder individually. However, this is not the case in U-Net. U-Net is an encoder-decoder network in which shallow layers are linked with deeper levels through skip connections. This implies there are no true 'encoder' and 'decoder' portions, in the sense of mapping the sample onto a well-defined latent space and computing the output from there. We cannot divide a U-Net into sections and utilise them independently since the input – and all its intermediate representations – are required to compute the output. Figure 3 shows the architecture of UNet.

UNet takes two different paths: one that contracts and one that expands. The conventional architecture is followed by the convolutional network's contracting path. It is made up of two repeated 3×3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a down-sampling 2×2 max pooling operation with stride 2. With each degree of down sampling, the number of feature channels doubles. At every step of the long voyage, the feature map is up-sampled. After that, a 2×2 convolutional ("up convolution") is used for half the features. Then a suitably trimmed shortened feature map from the decreasing path is concatenated with it. At the final layer, a 1×1 convolution is used to transform each 64-component feature vector to the appropriate number of classes.

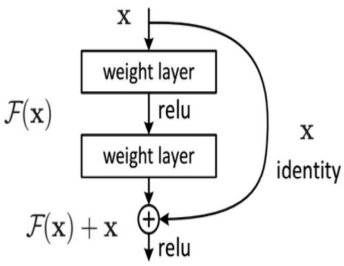
Figure 3 Architecture of UNet (see online version for colours)



3.3 ResUNet

More layers in conventional neural networks imply a better network. However, due to the vanishing gradient problem, the first layer's weights will not be updated successfully by back-propagation. The error gradient is minimal because it is back-propagated to prior layers via repeated multiplication. As a result, as the network grows in layers, its performance becomes saturated and quickly declines. ResUNet solves this problem by replacing UNet blocks with modified residual blocks. In this network, there are a series of stacked residual units. Figure 4 shows a single residual block used in ResUNet.

Figure 4 Single residual block



4 Experimentation

4.1 Dataset

The experimentation is performed on images from publically available datasets. The dataset used includes 208 satellite images of wildfires. The data is collected from varied resources available online. It is challenging to differentiate smoke from clouds since they look the same visually. Thus, to make the model more efficient in such circumstances, the dataset for training included images of fire, smoke, and clouds. Satellite pictures in the public domain were used for training and validation of data. Some sources include:

- NASA Earth Observatory – Home, 2021 (https://earthobservatory.nasa.gov/)
- GeoMAC Wildfire Application",Geomac.gov, 2021 (https://IoU.geomac.gov/)
- Australia's Deadly Wildfires in Photos: TheView from Space, 2021 (https://IoU.space.com/australia-wildfires-satellite- images- 2019-2020.html)
- National Aeronautics and SpaceAdministration, 2021 (https://IoU.nasa.gov/)

The images in the dataset are in RGB format. The images mainly include clouds, smoke and snow. Presence of smoke is considered as sign of fire. In this work, following types of data images are used for testing the algorithm:

1 images with no fire

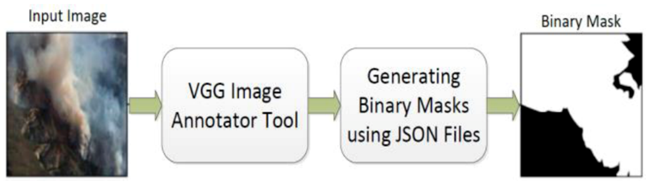
2 images with fire

3 images with smoke and clods

4 images with smoke and snow.

The VGG image annotator is used to annotate the dataset (https://IoU.robots.ox.ac.uk/~vgg/software/via/). Polyline marking is used to mark and annotate the region engulfed in smoke or fire and. After that, JSON files are created from the annotated images. The annotated areas are highlighted in white, while the unmarked areas are left in black. Figure 5 shows mask created for a sample image.

Figure 5 Ground truth mask for a sample image (see online version for colours)



4.2 Performance parameter

For quantitative analysis of the performance of the implemented methods, BCE Dice loss and IoU Score are used as performance measures.

4.2.1 BCE dice loss

This loss combines Dice loss and the conventional binary cross-entropy (BCE) loss, commonly used in segmentation models. Combining the two techniques allows for some loss of variety while making use of BCE's stability.

4.2.2 IoU score

The Jaccard index, commonly known as the Intersection over Union (IoU), is the most often used assessment measure for tasks including segmentation, object detection, and tracking. Intersection over Union is nothing more than a ratio. In this case, the numerator computes the area of overlap between the predicted and ground-truth bounding boxes. The area of union, or, to put it another way, the area covered by both the predicted and ground-truth bounding boxes, is the numerator. Our final score – the Intersection over Union – is calculated by dividing the overlap by the area of the union, and a score of 0.5 or more is considered acceptable.

5 Results and discussion

Quantitative and qualitative results obtained in this work are given in this section.

5.1 Training of the networks

The training is carried out on a Google Colab GPU, an Nvidia Tesla T4, with 16GB GPU Memory and up to 8.1 TFLOPS. TensorFlow is used to train UNet and ResUNet, while PyTorch with CUDA is utilised for training Mask RCNN.

UNet and ResUNet both required about 20 min to train. Both models were set to train for 60 epochs but were terminated before then, utilising TensorFlow's callback feature. The callback halted the training for various reasons, including preventing over fitting, failing to improve the loss for ten consecutive epochs, and so on. The Mask R-CNN model is trained for 1500 epochs.

The training process for UNet, ResUNet, and Mask R_CNN is summarised in Figure 6. It is observed that after 30 epochs, UNet's loss and IoU remain consistent, with 0.33 as the highest IoU score. ResUNet performs better and continued to learn, reaching the highest IoU score of 0.38. For the proposed Mask R-CNN technique, the IoU is roughly 0.925, and the total loss reduces over the training time. The computation time for the three algorithms is mentioned in Table 1.

Table 1 Computation time for training

UNet	*ResUNet*	*Mask RCNN*
17 min	24 min	12 min

The IoU score values obtained are compared with those obtained with existing methods. Table 2 shows the comparison of IoU values obtained.

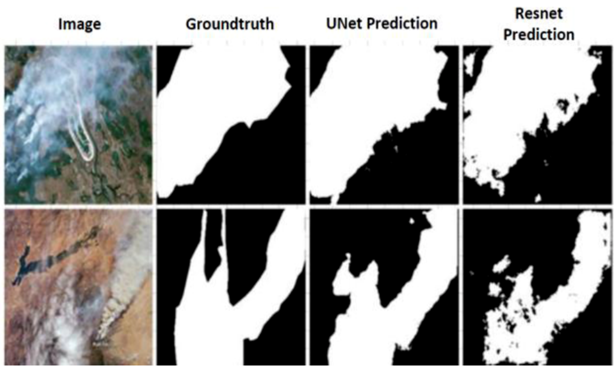
Table 2 Comparison of IoU scores

Approach	*IoU score*
FireDGWF (Pan et al., 2021)	0.69
Modified VGG16 (Frizzi et al., 2021)	0.86
Mask R-CNN (This work)	0.925

It is seen that the Region Based Segmentation approach yields a better result than Fully Convolutional Network-Based Semantic Segmentation. It is observed that the methods based on encoder-decoder models proposed in Pan et al. (2021) and Frizzi et al. (2021) give IoU values of 0.69 and 0.86, respectively. The Region-Based Segmentation approach using Mask RCNN is proven to beat these numbers by attaining an IoU of 0.925. The proposed novel approach can differentiate between multiple clouds of smoke and fires successfully.

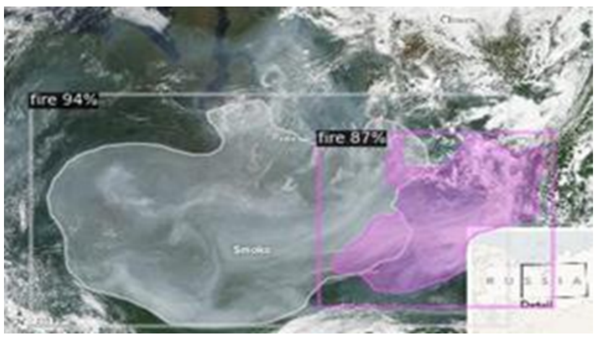
Qualitative results obtained are shown in Figure 6 and Figure 7. Figure 6 shows the prediction of fire achieved with UNet and ResUNet techniques on sample images. It also shows the ground truth images available for comparison.

Figure 6 Prediction by UNet and ResNet (see online version for colours)



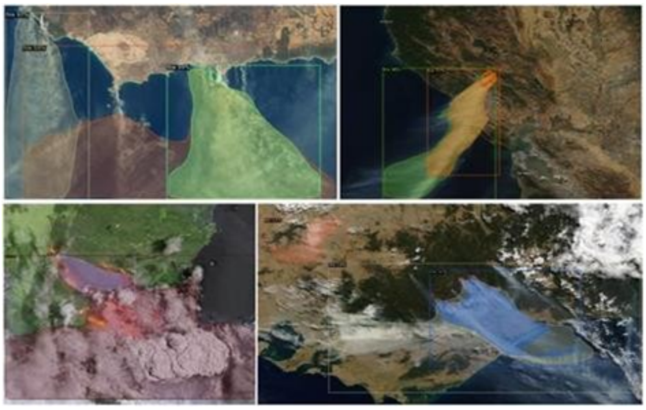
It can be seen that the prediction by the UNet technique is more precise as compared to the ResNet technique when compared with the ground truth. Both of these methods are not able to distinguish clouds and smoke successfully. Figure 7 shows the mask generated by the proposed Mask R-CNN method for an image having a smoke in snowy regions. UNet architecture gives good results for specific shapes of objects but fails when object have random shapes. As smoke takes random shapes, UNet fails to detect it efficiently.

Figure 7 Mask created by mask R-CNN method for snowy region (see online version for colours)



It is observed that the model can discern the differences between snow and smoke in icy regions. Figure 8 shows the mask generated by the proposed Mask R-CNN method for images having clouds and smoke. There is a good possibility that the model might confuse smoke and clouds due to similar colours and structure. Although, from the images in Figure 8, it is seen that, the proposed Mask R–CNN technique can distinguish between the clouds and smoke successfully.

Figure 8 Mask created by mask R-CNN method for cloudy image (see online version for colours)



6 Conclusion

In this paper, a semantic segmentation algorithm based on deep learning is proposed for early fire detection using satellite images. The existing methods using UNet and ResUNet are implemented for comparison purposes. It is observed that the proposed Mark R-CNN technique can distinguish between clouds and smoke and smoke and snow successfully. Mask R-CNN beats Res-UNet in terms of IoU score. Mask R-CNN has an IoU of 0.925, whereas UNet and Res-UNet have IoUs of 0.30 and 0.35, respectively. It can be concluded that the Region-Based Semantic Segmentation technique is more effective that the Fully Convolutional Network-Based Semantic Segmentation mechanism. Even though Mask R-CNN requires more epochs, converging to the local minimum takes less time than the other two techniques. Furthermore, the researchers can expand this study to incorporate night- time nano-satellite pictures. The limitation of this work is that in a few images, some of the background pixels get falsely detected as foreground pixels. This work can be taken ahead in the future by using an algorithm like Improved Mask R-CNN for fire detection. Also, the experimentation needs to be done on more samples from the various datasets.

References

Foivos, I.D., Waldner, F., Caccetta, P. and Wu, C. (2020) 'ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 162, pp.94–114, https://doi.org/10.1016/j.isprsjprs.2020.01.013

Frizzi, S., Bouchouicha, M., Ginoux, J-m., Moreau, E. and Sayadi, M. (2021) 'Convolutional neural network for smoke and fire semantic segmentation', *IET Image Processing, Institution of Engineering and Technology*, Vol. 15, No. 3, pp.634–647, https://doi.org/10.1049/ipr2.12046

Gagliardi, A., de Gioia, F. and Saponara, S. (2021) 'A real-time video smoke detection algorithm based on Kalman filter and CNN', *J. Real-time Image. Proc.*, https://doi.org/10.1007/s11554-021-01094-ioU

He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2020) 'Mask R-CNN', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 2, pp.386–397, doi: 10.1109/TPAMI.2018.2844175.

Huang, L., Liu, G., Wang, Y., Yuan, H. and Chen, T. (2022) 'Fire detection in video surveillances using convolutional neural networks and wavelet transform', *Engineering Applications of Artificial Intelligence*, Vol. 110, p.104737.

Kim, B. and Lee, J. (2019) 'A video-based fire detection using deep learning models', *Applied Sciences*, Vol. 9, No. 14, p.2862.

Krüll, W., Tobera, R., Willms, I., Essen, H. and von Wahl, N. (2012) 'Early forest fire detection and verification using optical smoke, gas and microwave sensors', *Procedia Engineering*, Vol. 45, pp.584–594, https://doi.org/10.1016/j.proeng.2012.08.208

Kumarguru, P.K. and Siau-Chuin, L.S. (2015) 'Fire Detection Algorithm Using Image Processing Techniques', International Conference on Artificial Intelligence and Computer Science [Internet], 2015 [cited 12 August, 2021], Available from: https://IoU.researchgate.net/publication/285580944_FIR E_DEtection_Algorithm_Using_Image_Processin G_Techniques

Li, P. and Zhao, W. (2020) 'Image fire detection algorithms based on convolutional neural networks', *Case Studies in Thermal Engineering*, Vol. 19, p.100625.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J. (2021) 'A survey of convolutional neural networks: analysis, applications, and prospects', *IEEE Transactions on Neural Networks and Learning Systems*.

Pan, H., Badawi, D., Zhang, X. and Cetin, A. (2020) 'Additive neural network for forest fire detection', *Signal, Image and Video Processing*, Vol. 14, pp.675–682.

Pan, J., Ou, X. and Xu, L. (2021) 'A collaborative region detection and grading framework for forest fire smoke using weakly supervised fine segmentation and lightweight Faster-RCNN', *Forests*, Vol. 12, p.768, https://doi.org/10.3390/f12060768

Panagiotis, B., Papaioannou, P., Dimitropoulos, K. and Grammalidis, N. (2020) 'A review on early forest fire detection systems using optical remote sensing', *Sensors*, Vol. 20, No. 22, p.6442, https://doi.org/10.3390/s20226442

Robinne, F-N. (2021) 'Impacts of disasters on forests', *Particular Forest Fires*, August, https://IoU.researchgate.net/publication/350850462_UNF F16

Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-net: convolutional networks for biomedical image segmentation', *Lecture Notes in Computer Science*, Vol. 9351, pp.234–241, https://arxiv.org/abs/1505.04597

Rui, B.R., Chen, C.C., Jing, Y.J., Weiguo, S.I.O.U. and Siuming, L.S. (2019) 'SmokeNet: satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention', *Remote Sensing*, Vol. 11, No. 14, p.1702, https://doi.org/10.3390/rs1114170

Saeed, F., Paul, A., Kumar, K. and Nayyar, A. (2020) 'Convolutional neural network based early fire detection', *Multimedia Tools and Applications*, Vol. 79, pp.9083–9099.

Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X. and Yang, H. (2022) 'Crack detection and comparison study based on faster R-CNN and mask R-CNN', *Sensors*, Vol. 22, No. 3, p.1215.

Zhang, Z., Liu, Q. and Wang, Y. (2018) 'Road extraction by deep residual U-net', *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp.749–753, https://doi.org/10.1109/LGRS.2018.2802944

Websites

Christine Lunsford, Australia's Deadly Wildfires in Photos: The View from Space [Internet], Space.com, 2021, https://IoU.space.com/australia-wildfires-satellite-images-2019-2020.html

GeoMAC Wildfire Application [Internet], Geomac.gov, 2021, Available from: https://IoU.geomac.gov/

NASA Earth Observatory – Home [Internet], Earthobservatory.nasa.gov, 2021 [cited 12 August 2021], Available from: https://earthobservatory.nasa.gov/.

National Aeronautics and Space Administration [Internet], NASA, 2021 [cited 12 August, 2021], Available from: https://IoU.nasa.gov/

Visual Geometry Group – University of Oxford [Internet], Robots.ox.ac.uk, 2021 [cited 12 August 2021], Available from: https://IoU.robots.ox.ac.uk/~vgg/software/via/