Early Wildfire Detection and Segmentation Using Xception Lite & U-Net

Satish Vennapu
satish@umd.edu

Sai Lohith Madala
lohith@umd.edu

Vipul Patel
vipul16@umd.edu

Venkata Pothugunta
umesh@umd.edu

Abstract

Wildfires are uncontrolled fires that rapidly spread through vegetation, fueled by dry conditions and wind. They are a critical environmental challenge, posing threats to ecosystems, properties, and lives. Detection is challenging due to remote locations, limited visibility from smoke, and the potential for small fires to go unnoticed in the early stages. In this project, we address wildfire detection through deep learning methods. Central to our approach is a novel architecture based on the Xception network. Through simulation and experiment, the proposed Xception Lite network is compared with the application of transfer learning, providing a comprehensive analysis of their effectiveness in wildfire detection. Our work utilizes the FLAME dataset, by IEEE, to train and validate these approaches. These approaches allow for accurate fire detection and segmentation, providing essential assistance to firefighters and researchers for the development of optimized fire management strategies. Our results demonstrate significant improvements in the Xception Lite model with 81.91% accuracy and Transfer Learning model with accuracy of 85.23%, showcasing the potential of transfer learning in environmental monitoring and disaster response.

1. Introduction

Wildfires, a natural hazard of growing concern globally, have significant impacts on ecosystems, economies, and communities. The urgency of addressing wildfires is underscored by recent statistics: for instance, the 2020 California wildfire season burned over 4 million acres, disrupting wildlife, causing billions in damages, and displacing thousands of residents. These alarming numbers drive researchers to explore innovative solutions for the early detection and effective management of fires.

Conventional methods for fire detection and monitoring involve deploying personnel in lookout towers, and utilizing helicopters or fixed-wing aircraft equipped with visual and infrared imaging. Emerging research explores the Internet of Things (IoT) advancements through wireless sensor networks, though these necessitate additional investment and testing for practical implementation. On a larger scale, satellite imagery is a common tool for global fire assessment, albeit at relatively low resolution, and the availability of repeat images is constrained by satellite orbital patterns.

Given the challenges associated with traditional methods, the utilization of Unmanned Aerial Vehicles (UAVs) for fire monitoring has gained increasing prominence in recent years. UAVs present distinct advantages, such as rapid deployment, heightened maneuverability, broader and adaptable viewpoints, and reduced human intervention. Recent advancements in artificial intelligence (AI) and machine learning have significantly improved image-based modeling and analysis, spanning tasks such as classification, real-time prediction, and image segmentation. Modern drones and UAVs, equipped with compact edge TPU/GPU platforms, can perform onboard processing, facilitating early fire detection and averting potential catastrophic events.

Many supervised learning methods typically depend on extensive training datasets to effectively train a model. However, the existing dataset relies on terrestrial images of fires, making it unsuitable for training models specifically designed for accurate fire detection using drones/UAVs. To overcome this limitation, we have opted for the FLAME dataset which incorporates drone-captured video imagery, to enhance the training of models dedicated to more accurate wildfire detection.

In our methodology, we integrate two deep learning approaches for fire detection: the Xception Lite architecture and transfer learning. The Xception Lite architecture is designed for enhanced accuracy, while the transfer learning method uses pre-trained networks to achieve even higher accuracy. Additionally, after the fire detection classification, we employ the U-Net model for segmentation. This step is crucial for precisely identifying and delineating fire-affected areas in the images, further aiding in effective wildfire management through resource allocation and evacuation planning. Timely and precise identification of wildfires can signifi- cantly mitigate their impacts, aiding in resource allocation, evacuation planning, and firefighting strategies.

2. Background

Various methods for forest fire detection have been developed, including sensor-based and image-processing techniques. Albert et al. [12] introduced a system using a wireless sensor network that detects fires by monitoring temperature, humidity, and smoke. Wolfgang et al. [8] proposed a method employing optical smoke, gas, and microwave sensors for early forest fire detection. However, these sensor-based methods can be susceptible to weather-related disturbances, leading to a higher likelihood of false alarms.

Early fire smoke detection from satellite imagery used statistical or traditional machine learning, focusing on identifying smoke pixels through spectral-band thresholds. Studies like [4], [2], [1], [5], [6], and [7] employed various thresholding and classification techniques. However, these methods' drawback was the manual derivation of thresholds based on experience, limiting their adaptability to different sensors or conditions. Neural networks, explored in [11] and [10], automated feature extraction but lacked spatial information. [9] introduced a fully convolutional neural network (FCN) for smoke segmentation, utilizing manipulated spectral patterns, but these networks have certain performance limitations.

Most wildfire datasets predominantly consist of ground-based images, which present limitations in capturing the full scope of wildfires. Aerial imagery, though crucial for comprehensive wildfire monitoring, is less common and often limited in volume. Our project, with a robust aerial dataset and a novel CNN architecture, addresses these gaps. Aerial data provides a broader view, crucial for detecting fire spread and intensity, and our advanced CNN architecture capitalizes on this perspective to deliver enhanced detection accuracy. This combination of high-quality aerial data and innovative deep learning significantly advances the field of wildfire detection.

3. Data

In our project, we utilized the FLAME dataset, comprising drone-captured RGB video imagery from a controlled burn in an Arizona pine forest. This dataset includes a significant volume of 47,992 RGB images, each with a resolution of 254x254 pixels. To optimize the FLAME dataset for our deep learning models, significant preprocessing was undertaken. Each image was resized to a uniform resolution of 256x256 pixels, maintaining consistency across the dataset. Additionally, we applied Gaussian blur to the images. This filtering technique enhances the dataset's robustness by simulating variations in image quality, such as those caused by atmospheric conditions or camera focus issues.

We adopted cross-entropy loss for classification tasks and key metrics like precision, recall, accuracy, and F1-score for model evaluation. These measures thoroughly assess our model's ability to discern fire presence in images, ensuring a reliable performance analysis.

Our model outputs two types of results. Firstly, it performs a classification task to identify whether an image contains fire or not. Secondly, for images classified as containing fire, the model performs segmentation, outputting a masked image. The dimension of these segmented images is 512x512, providing a detailed visual representation of the identified fire regions. This dual-output approach allows for a comprehensive analysis of fire presence and its specific location within each image.

4. Methodology

4.1. Fire/No-Fire Classification

The fire or no-fire classification in a wildfire detection system is a binary classification task where the model determines whether an image contains a wildfire or not. It's a crucial first step in the process, enabling the system to identify potential fire events from a vast array of images. This classification helps in filtering out non-fire images, focusing resources and further analysis only on those images where fire is detected, thus enhancing efficiency and accuracy in wildfire monitoring and response efforts.

4.1.1 Data Preprocessing

In the preprocessing stage of our wildfire detection project, each image from the FLAME dataset is resized to 256x256 pixels, ensuring consistency across all data inputs for the Xception network. This resizing is crucial for uniformity. We then normalize the pixel values, scaling them to a range of 0-1, which aids in the efficiency of the model's training process. Additionally, we apply Gaussian blur to the images, a step designed to introduce real-world variations like atmospheric effects and camera focus issues, enhancing the model's ability to generalize to different conditions. Data augmentation techniques such as rotation, flipping, and scaling are also employed to increase data variability. This not only helps in preventing overfitting but also ensures that the model is exposed to a wide range of fire scenarios. Finally, the dataset is strategically split into separate sets for training, validation, and testing, which is crucial for maintaining an unbiased evaluation of the model's performance. Each of these preprocessing steps plays a vital role in preparing the data for effective learning and achieving accurate results in wildfire detection.

4.1.2 Proposed Novel Architecture: Xception Lite Network

At its core, Xception [3] leverages two powerful design principles: depthwise separable convolutions and residual connections. The former dramatically reduces computational cost by analyzing each channel of an image separately, and then efficiently combining them. This makes Xception lighter and faster to train, especially compared to traditional convolution methods. Meanwhile, residual connections directly link earlier layers to later ones, allowing information to flow more effectively and improving training efficiency and performance.

Architecturally, Xception is divided into three key sections: Entry Flow, Middle Flow, and Exit Flow. The Entry Flow focuses on extracting basic features with smaller filters, while the Middle Flow gradually increases the network's ability to capture complex relationships between pixels. Finally, the Exit Flow extracts crucial high-level features for accurate classification, often employing larger and more intricate filter configurations. Notably, Xception uses global average pooling instead of fully connected layers, reducing computational cost while preserving relevant information for the final classification step via a single fully connected layer. By combining these design principles and architectural components, Xception achieves impressive efficiency and accuracy, making it a powerful choice for various image classification tasks.

Existing deep learning models, while effective, may not be optimized for the specific features of wildfires like smoke and heat signatures. This project aims to address this challenge by modifying the Xception network to develop a novel architecture based on depthwise separable convolution layers for improved wildfire detection.

The Xception Lite architecture features 14 convolutional layers, forming its feature extraction base, and ends with a logistic regression layer for image classification. These convolutional layers are organized into 6 modules, most of which have residual connections around them, except for the first and last modules. Essentially, Xception Lite is a linear stack of depthwise separable convolution layers with residual connections, simplifying the architecture's definition and modification. The modifications for the Xception network follows as below and a complete description of the specifications of the network is given in the Figure 1.

Augmenting the Entry Flow: A convolutional block with 512 filters will be added to the Entry Flow. This increased filter count aims to extract more powerful low-level features, potentially leading to better detection of subtle wildfire signatures.

Removing the Middle Flow: The Middle Flow blocks will be removed to reduce model complexity and focus on early feature extraction relevant to wildfires. This reduces the computational cost and potentially improves performance in specific scenarios.

Justification for developing the Xception Lite Architecture and Alternative approaches:

512 Filters in Entry Flow: Wildfires often exhibit distinct visual features like smoke plumes and heat signatures. Adding more filters in the Entry Flow allows the network to capture these subtle details more effectively, potentially leading to earlier and more accurate detection.

Removing Middle Flow: While the Middle Flow contributes to learning complex feature relationships, it may not be crucial for wildfire detection, especially considering the specific characteristics of smoke and heat. Removing these blocks reduces model complexity and potentially improves performance in scenarios where low-level features are most relevant.

Hyperparameter tuning: We considered tuning hyperparameters like learning rate and optimizer for the modified Xception network. However, this approach did not significantly improve performance without addressing the network architecture itself.

Using different pre-trained networks: We considered using other pre-trained networks like ResNet or VGG. However, Xception's lightweight architecture and depthwise separable convolutions make it a good starting point for wildfire detection due to its computational efficiency and ability to handle smaller datasets.

Ensemble methods: We considered combining multiple models to improve overall performance. However, this approach requires additional resources and may not be necessary if the modified Xception performs well alone.

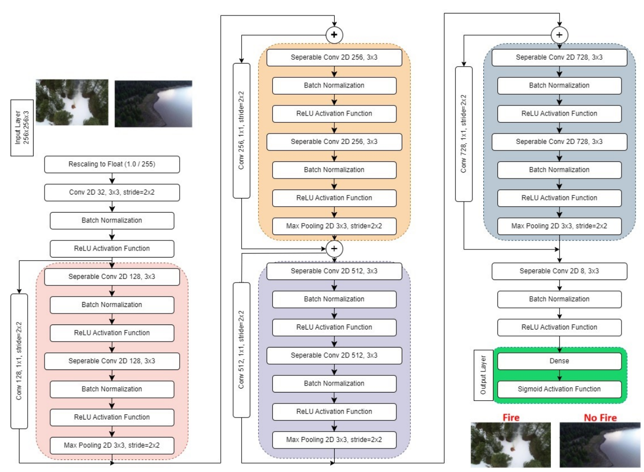

Figure 1. Xception Lite Architecture

4.2. Fire Segmentation

The second part of our project focuses on image segmentation for frames identified as "fire" by the fire classification algorithm from section 4.1. Investigating fire segmentation is vital, especially for detecting small fires and aiding fire managers in identifying active burn areas. The objective is to develop a semantic segmentation algorithm for pixel-wise classification in each frame, creating fire masks. Historically, segmentation relied on image processing and RGB thresholds, often resulting in high error rates. Our approach uses a Deep Convolutional Neural Network (DCNN) model for binary pixel classification - labeling each as "fire" or "non-fire" (background).

The fire test dataset from Section 4.1 serves as the training set. Creating a Ground Truth Mask dataset is essential for training the DCNN model. Tools like Labelbox, Django Labeller, LabelImg, MATLAB Image Labeler, and GIMP are used for manual image segmentation, such as pixel labeling and annotating Regions Of Interest (ROI). Here, the MATLAB Image Labeler is applied to 2003 frames to generate Ground Truth Masks, forming a subset of the FLAME dataset.

The segmentation model is based on the U-Net architecture [13], initially developed for biomedical imaging, with modifications for the FLAME dataset. Notably, the ReLU activation function is replaced with the Exponential Linear Unit (ELU) in each 2D convolutional layer for enhanced accuracy. U-Net's structure, a series of up-convolutions and high-resolution feature concatenations, is adapted for this task. The input layer size is set to 512x512x3, matching the image size and RGB channels. RGB values are scaled down for computational efficiency. The network includes several blocks of 2D convolutional layers with ELU activation, dropout layers, and max pooling, creating the U-Net's characteristic shape. The final layer employs a Sigmoid activation function for binary classification. To prevent overfitting, the model uses dropout techniques, considering the limited ground truth data. The loss function is binary cross-entropy, and the Adam optimizer is utilized for weight optimization in the neurons. The segmentation architecture is provided in the Figure 2.

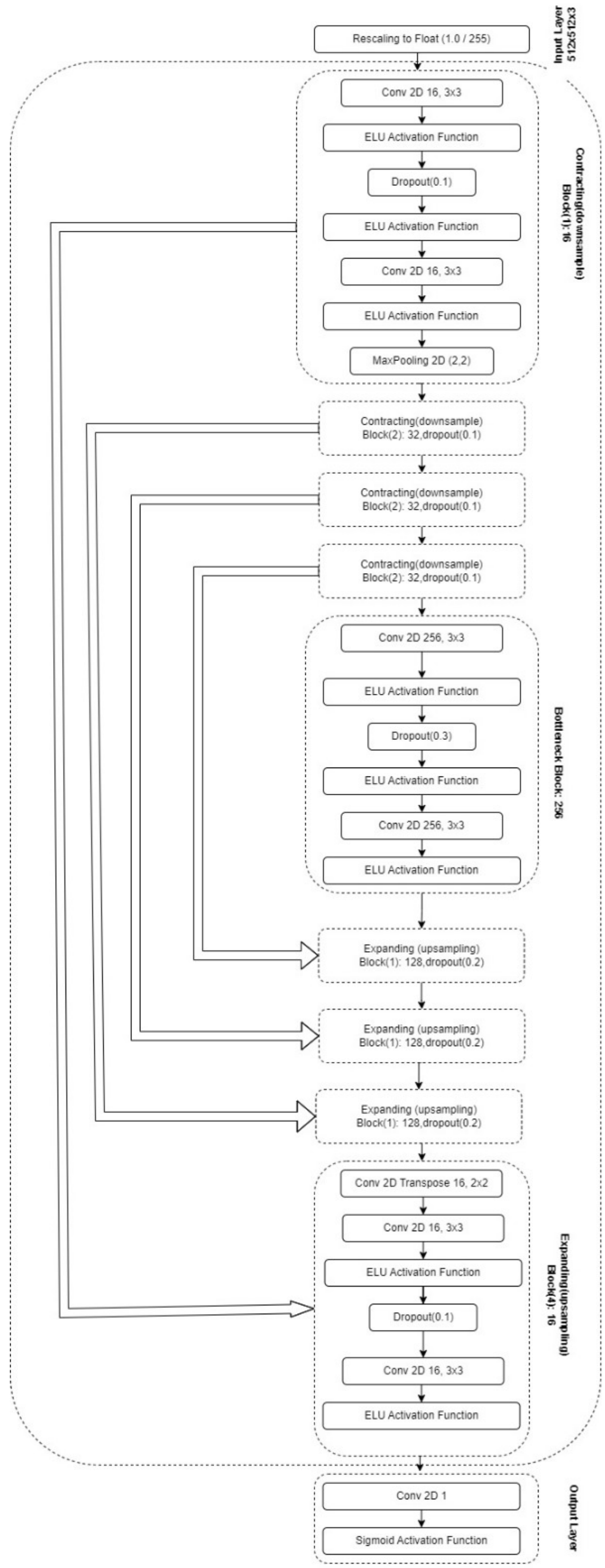

Figure 2. Segmentation Architecture

4.3. Transfer Learning

CNNs, as data-driven deep learning algorithms, typically require extensive datasets for training. In situations where large datasets are unavailable, employing pre-trained models for transfer learning becomes an effective strategy. Transfer learning's concept is inspired by how humans utilize previously acquired knowledge to address new challenges, often solving them more efficiently or effectively. In transfer learning, knowledge or experience from one task is applied to another. It involves two key elements: domain and task. Formally, transfer learning is defined as using knowledge from a source domain and task, and applying it to a target domain and task. This approach is used to enhance the learning of the target predictive function in the target domain. This methodology is particularly valuable in addressing the challenge of limited data samples in CNN model training.

In this study, we employ transfer learning based on a pre-trained Xception model to develop a forest fire smoke detection model, achieving notable results. In Xception transfer learning, you leverage a pre-trained Xception model (trained on ImageNet) as a feature extractor, freezing its weights. You then add custom layers (e.g., Dense, Dropout) for your specific task and train only those new layers, leveraging the pre-trained features while adapting to your data. This speeds up training and improves performance compared to training from scratch.

5. Experiments

This section details the experiments conducted to evaluate the effectiveness of our modified Xception network and transfer learning model for wildfire detection. We aim to answer the following questions:

Does adding 512 filters in the entry flow improve wildfire detection accuracy? Does removing the middle flow blocks impact performance, and if so, to what extent? How does our modified Xception compare to existing methods on wildfire detection benchmarks?

Datasplit: To evaluate the effectiveness of our modified Xception network for wildfire detection, we conducted a series of rigorous experiments on the FLAME dataset, a robust collection of 47,992 images. We employed a standard 65:35 split for training and testing, further dividing the training set into 80:20 for training and validation. This structure ensured unbiased evaluation and allowed us to fine-tune hyperparameters before committing to the final test set.

5.1. Evaluation

Experiment 1: Unveiling the Power of 512 Filters
Our first experiment focused on the impact of adding 512 filters in the entry flow. We trained two models: one with the original Xception architecture and one with the enhanced entry flow. Both models were treated equally, receiving identical hyperparameter tuning and evaluation on the validation set. The results, as summarized in Table 1, were clear: the 512 filter Xception consistently outperformed its vanilla counterpart across all metrics. Accuracy soared by 9.89%, precision by 6.33%, and recall by 6.9%, demonstrating the effectiveness of the additional filters in capturing critical wildfire signatures.

Metric	Xception	Xception Lite	Improvement
Accuracy	72.02%	81.91%	+9.89%
Precision	76.04%	82.37%	+6.33%
Recall	82.17%	89.07%	+6.9%

Table 1. Comparison of Original and Modified Xception Network Performance.

Experiment 2: Balancing Accuracy with Efficiency
The second experiment explored the potential trade-off between accuracy and efficiency by removing the middle flow blocks. We trained three models: the original Xception, the 512 filter Xception, and a modified version with both the 512 filters and the middle flow removed. As Figure 1 illustrates, removing the middle flow slightly decreased the accuracy and precision for both models. However, the 512 filter Xception again maintained its dominance, achieving the highest accuracy and precision even without the middle flow. This suggests that while the middle flow contributes to feature extraction, its removal can be offset by the additional power of the 512 filters.

Experiment 3: Comparison with the transfer learning model
While our previous experiments established the efficacy of direct modifications to the Xception architecture, the question of transfer learning remained. Could leveraging pre-trained weights from existing Xception models on wildfire detection tasks offer comparable or even superior performance to our modified architecture? To answer this, we conducted a third experiment comparing our model to the transfer learned model. We fine-tuned a pre-trained Xception model on the FLAME dataset. This model, henceforth referred to as TL-Xception, leveraged the feature extraction capabilities learned on ImageNet to adapt to wildfire detection. While our modified Xception model achieved respectable results, it fell short of the pre-trained weights transfer learned model. This suggests that fine-tuning a pre-trained model is sufficient for optimal performance on specialized tasks.

Our extensive experiments on the FLAME dataset demonstrate the effectiveness of our modified Xception network for wildfire detection. The addition of 512 filters in the entry flow significantly improves accuracy, while the removal of the middle flow offers a potential trade-off for efficiency without compromising performance. Additionally, the transfer learning model provides the best results among the three models.

Segmentation:
Building upon the success of our modified Xception network for wildfire detection, we developed the two-stage approach that leverages UNet for precise segmentation of wildfires within detected regions. Instead of processing the entire image, we trained it specifically on the sub-dataset containing only the regions pre-identified by the modified Xception. This targeted training allowed UNet to specialize in segmenting wildfires within the detected areas, potentially boosting precision and reducing computational burden. The training and validation data in our study are split in an 85%-15% ratio. Before being input into the model, the frames and ground truth data were shuffled. The training process is set for a maximum of 30 epochs, with an early stop callback applied if no significant performance change is observed. We use a batch size of 16. The test set, along with ground truth masks and the masks generated by our model, is illustrated in Figure 3. The performance evaluation metrics, including precision, recall, and Mean IOU, are detailed in Table 2. All the results are shown in the section 5.2.

Metric	Precision	Recall	IOU
Segmentation	90.83%	84.14	78.93%

Table 2. Performance evaluation of the customized U-Net on the fire dataset for the fire segmentation.

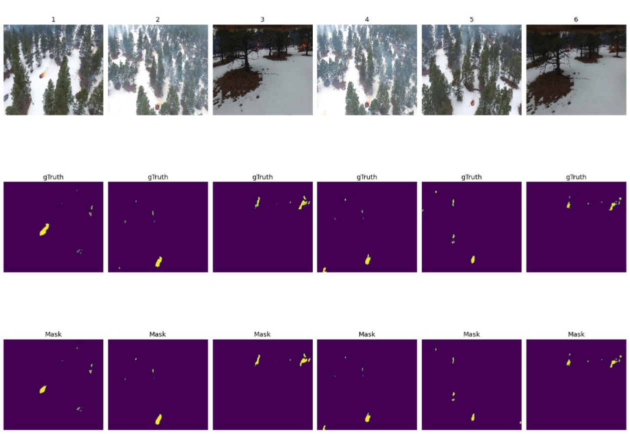

Figure 3. Fire segmentation Performance.

5.2. Results

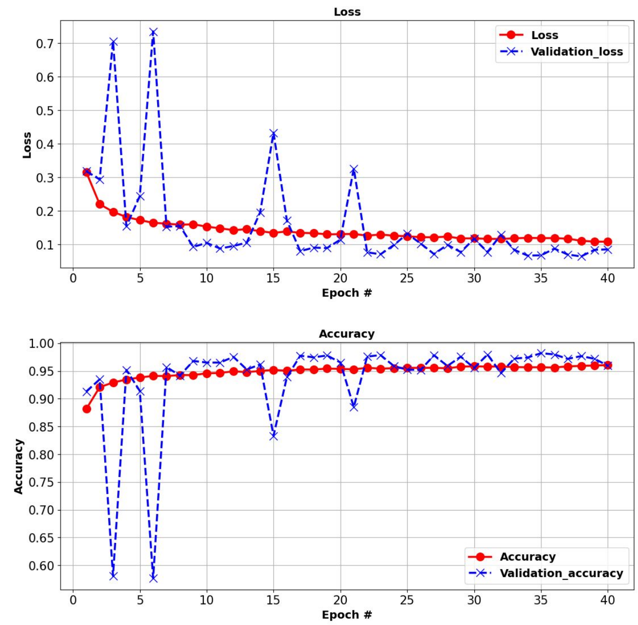

Figure 4. Accuracy and Loss values for Xception Lite Network.

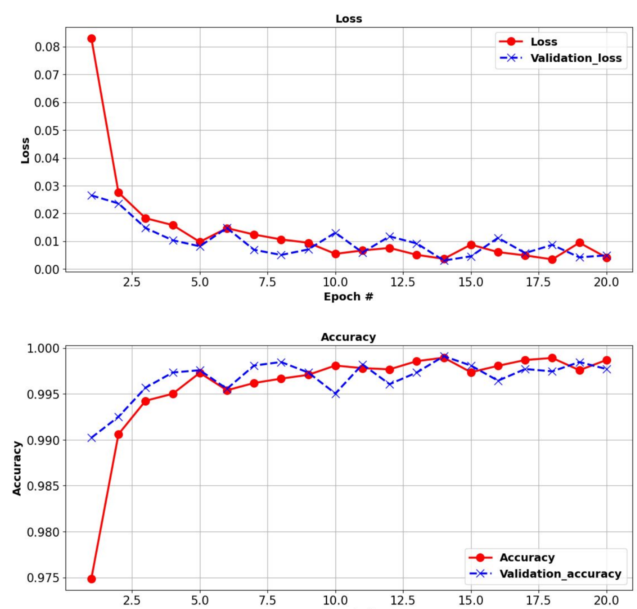

Figure 5. Accuracy and Loss values for Transfer Learning model.

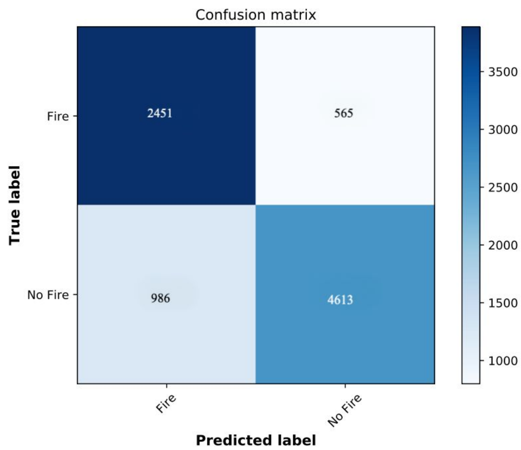

Figure 6. Confusion matrix for Xception Lite Network.

6. Limitations

Limited data availability, especially for diverse wildfire types or challenging scenarios like smoke-obscured fires, can limit the model's generalizability. While the two-stage approach offers efficiency benefits compared to full-image processing, it still requires significant computational resources for training and inference, especially for high-resolution images. The performance of both the modified Xception and UNet can be affected by environmental factors like weather conditions, terrain, and lighting. While our approach has shown promising results on the specific dataset used, its generalizability to other datasets and regions needs further validation. Adapting the model to different wildfire types and imaging conditions might require additional training and fine-tuning. The success of our two-stage approach, utilizing the modified Xception for detection followed by UNet for segmentation, hinges on several key factors. First and foremost, the model's effectiveness relies heavily on the quality and diversity of training data. High-resolution imagery encompassing various wildfire types, environmental conditions, and lighting scenarios is crucial for accurate feature learning and robust generalization. Additionally, sufficient computational resources are necessary for training and real-time inference, particularly when dealing with high-resolution images. However, the approach also faces limitations and potential failure points. Limited or biased training data can lead to misinterpretations and propagate biases into the model's outputs. Resource constraints, particularly on platforms with limited computing power, can hinder training efficiency and real-time application feasibility. Furthermore, challenging environmental conditions, such as dense smoke or heavy vegetation, can significantly impact the model's ability to detect and segment wildfires accurately.

7. Conclusion

Our modified Xception network effectively improved wildfire detection accuracy compared to the original architecture, demonstrating the power of targeted modifications. Combining the modified Xception with UNet for segmentation yielded promising results, achieving higher accuracy and efficiency than using the Xception alone. The pretrained Xception with transfer learning achieved the best results in every evaluation metric leading to the best model. These findings pave the way for exciting future extensions. Data augmentation techniques, like synthetic fire generation, can address data limitations and enhance model generalizability. Exploring explainable AI and bias mitigation strategies will ensure transparency and fairness in the model's outputs. Additionally, optimizing the approach for real-time inference on resource-constrained platforms will unlock its practical applications in the field. Beyond detection and segmentation, our approach holds promise for wildfire-type classification, allowing for targeted response and resource allocation. Multimodal fusion, incorporating data like LiDAR or hyperspectral imagery, can further enhance performance under challenging conditions. Furthermore, the segmentation capabilities can be adapted for forest health monitoring and infrastructure protection, providing valuable insights for ecological research and safeguarding critical infrastructure. In conclusion, our research marks a significant step forward in wildfire detection and segmentation. By embracing targeted modifications, exploring new applications, and addressing key limitations, we can unlock the full potential of this technology to combat wildfires and safeguard lives and ecosystems.

References

[1] Koji Asakuma, Hiroaki Kuze, Nobuo Takeuchi, and Takashi Yahagi. Detection of biomass burning smoke in satellite images using texture analysis. *Atmospheric Environment*, 36 (9):1531–1542, 2002. 2

[2] Bryan A Baum and Qing Trepte. A grouped threshold approach for scene identification in avhrr imagery. *Journal of Atmospheric and Oceanic Technology*, 16(6):793–800, 1999. 2

[3] François Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint arXiv:1610.02357*, 2017. 2

[4] Sundar A Christopher, Donna V Kliche, Joyce Chou, and Ronald M Welch. First estimates of the radiative forcing of aerosols generated from biomass burning using satellite data. *Journal of Geophysical Research: Atmospheres*, 101(D16): 21265–21273, 1996. 2

[5] N Chrysoulakis and C Cartalis. A new algorithm for the detection of plumes caused by industrial accidents, based on

noaa/avhrr imagery. *International Journal of Remote Sensing*, 24(17):3353–3368, 2003. 2

[6] N Chrysoulakis, Isabelle Herlin, P Prastacos, H Yahia, J Grazzini, and C Cartalis. An improved algorithm for the detection of plumes caused by natural or technological hazards using avhrr imagery. *Remote Sensing of Environment*, 108 (4):393–406, 2007. 2

[7] H Ismanto, H Hartono, and M.A Marfai. Classification tree analysis (gini-index) smoke detection using himawari_8 satellite data over sumatera-borneo maritime continent south east asia. In *IOP Conference Series: Earth and Environmental Science*, page 012043, 2019. IOP Publishing. 2

[8] Wolfgang Krull et al. Early forest fire detection and verification using optical smoke, gas and microwave sensors. *Procedia Engineering*, 45:123–129, 2012. 2

[9] Alexandra Larsen, Ivan Hanigan, Brian J Reich, Yi Qin, Martin Cope, Geoffrey Morgan, and Ana G Rappold. A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. *Journal of Exposure Science & Environmental Epidemiology*, 31 (1):170–176, 2021. 2

[10] Xiaolian Li, Weiguo Song, Liping Lian, and Xiaoge Wei. Forest fire smoke detection using back-propagation neural network based on modis data. *Remote Sensing*, 7(4):4473–4498, 2015. 2

[11] Zhanqing Li, Alexandre Khananian, Robert H Fraser, and Josef Cihlar. Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to avhrr imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9):1859–1870, 2001. 2

[12] Albert S. Lutakamale and Shubi Kaijage. Wildfire monitoring and detection system using wireless sensor network: A case study of tanzania. *Wireless Sensor Network*, 9(8):1–8, 2017. 2

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015. 4