



Research paper

Wildfire and smoke early detection for drone applications: A light-weight deep learning approach

Abhinav Kumar, Adolfo Perrusquía*, Saba Al-Rubaye, Weisi Guo

School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK



ARTICLE INFO

Keywords:

Deeplabv3+
Mobile vision transformers
Mobilenet
Wildfire and smoke detection
Segmentation

ABSTRACT

Drones have become a crucial element in current wildfire and smoke detection applications. Several deep learning architectures have been developed to detect fire and smoke using either colour-based methodologies or semantic segmentation techniques with impressive results. However, the computational demands of these models reduce their usability on memory-restricted devices such as drones. To overcome this memory constraint whilst maintaining the high detection capabilities of deep learning models, this paper proposes two lightweight architectures for fire and smoke detection in forest environments. The approaches use the Deeplabv3+ architecture for image segmentation as baseline. The novelty lies in the incorporation of vision transformers and a lightweight convolutional neural network architecture that heavily reduces the model complexity, whilst maintaining state-of-the-art performance. Two datasets for fire and smoke segmentation, based on the Corsican, FLAME, SMOKE5K, and AI-For-Mankind datasets, are created to cover different real-world scenarios of wildfire to produce models with better detection capabilities. Experiments are conducted to show the benefits of the proposed approach and its relevance in current drone-based wildfire detection applications.

1. Introduction

A wildfire refers to an unregulated blaze that swiftly expands across vegetative ecosystems and natural terrains, encompassing areas like woodlands, grassy fields, or bushlands (El Debeiki et al., 2024). Wildfires can occur naturally due to lightning strikes, or they can be caused by human activities, such as campfires, arson, discarded cigarettes, or sparks from machinery. The severity and magnitude of spread of a wildfire depends on several factors such as availability of suitable fuel, favourable temperatures, climate, wind, and terrain (Pausas and Keeley, 2021).

Recently, unmanned aerial vehicle (UAV) platforms also known as drones have found widespread applications across various sectors that require monitoring from a distant height because of their distinct characteristics such as the ability to fly at low altitudes, low cost, and high flexibility (Yao et al., 2019; Gruffeille et al., 2024; Lakhan et al., 2022). Since wildfires are often spread over large areas and are unpredictable and dynamic in nature, the role of drones in their early detection is crucial (Akhoulfi et al., 2021).

Various forest fire detection methods using drones equipped with deep learning techniques were reviewed in Yandouzi et al. (2022b). They found that deep learning-based methods performed better compared to traditional sensing methods. They concluded that drones could

provide accurate and real-time data on the location, direction, and intensity of fire. To achieve this, drones need to be equipped with advanced sensors and imaging technologies (Perrusquía et al., 2024). With these enhancements, drones become effective tools for gathering crucial information during wildfires. The challenge of integrating novel deep learning technologies into drones is exasperated by its limited memory processing. This means that we cannot use deep models with high number of parameters because the memory cannot deal with this high-demand computation. Therefore, the development of new approaches that complies with the memory constraints, whilst achieving high-accurate fire detection results is required.

Hence, the main problem we aim to solve in this paper is related to the design of a drone-compatible lightweight deep learning model architecture for fire and smoke detection that achieves high detection results. To this end, we will combine the benefits of vision transformers, and autoencoder architectures to create a lightweight deep learning model that achieves competitive results, whilst considering the memory constraints of drones.

The outline of the paper is as follows: Section 2 summarizes the related work and state-of-the-art. Section 3 describes the proposed methodology for fire and smoke segmentation. Section 4 provides the datasets and loss functions used in this work. Section 5 discuss the

* Corresponding author.

E-mail address: adolfo.perrusquia-guzman@cranfield.ac.uk (A. Perrusquía).

experiments and results. Conclusions and final remarks are given in Section 6.

2. Related work

Methods for detecting wildfires and smoke using drones can be broadly categorized into two primary groups: conventional image processing methods based on predefined colour-based rules, and methods based on deep learning techniques.

Wildfire detection based in colour-based approach uses expert rules to extract fire pixels based in the saturation of the red colour in a RGB space (Chen et al., 2004; Abdusalomov et al., 2023). These approaches are quite sensitive to the diverse colour spectrum in imagery data and illumination conditions, e.g., sunshine. The YCbCr colour space (Wu and Zhang, 2018; Celik and Demirel, 2009) has been used to increase the robustness of colour-based approaches by incorporating the luminance and chrominance of the image as additional element in the design of expert rules. However, the presence of false positives is latent in this approach for pixels in the image that are on the borderlines of the established thresholds.

Deep learning segmentation methods have been used to identify fire including its shape and texture. Here, the most used approaches are based on convolutional neural networks (CNNs) (Lateef and Ruichek, 2019) such as U-Net (Ronneberger et al., 2015), Mask R-CNN (He et al., 2017), and deeplabv3 (Chen et al., 2019) architectures. These models are commonly used with colour-based approaches for semantic segmentation of fire pixels. Some of the approaches used in the literature for wildfire and smoke detection are given in the following section, highlighting their advantages and disadvantages and how the proposed approach deal with them.

2.1. Classification-based methods

In the context of wildfire and smoke detection, classification involves determining whether an image or video frame contains a fire or smoke, by assigning a label to the whole image. A time-efficient fire detection system for fire classification was applied by leveraging CNN and transfer learning (Hosni Mahmoud et al., 2022). In Chung et al. (2019) a combination of modified CNN architecture, VGGNet with colour space conversion, image filtering and other region-of-interest methodologies were applied to detect wildfire in domestic environments. Transfer learning techniques using multiple variants of pre-trained CNN architectures were used for the wildfire classification tasks (Yandouzi et al., 2022a). An artificial neural network approach was introduced to determine the presence of fire (Shamsoshoara et al., 2021) using a U-Net (Zunair and Hamza, 2021). Here, the approach segments the fire masks from video frames for binary classification. A unique Learning without Forgetting (LwF) approach was proposed to construct a model capable of detecting fire/smoke by leveraging pre-trained CNN architectures (Sathishkumar et al., 2023). All these techniques could accurately classify the presence of fire/smoke within an image, yet they fell short to pinpoint its location or delineate the expanse; they merely confirmed its existence somewhere in the frame.

2.2. Object detection-based methods

Object detection is computationally intensive than classification, but it provides more detailed information of the fire location. Faster R-CNN (Chollet, 2017), variants of YOLOv3 (Abdusalomov et al., 2021; Redmon and Farhadi, 2017), and SSD (Liu et al., 2016) were used for the identification of fire and smoke. The potential fire areas were identified in a frame based on their spatial characteristics using the Faster RCNN technique (Kim and Lee, 2019). In Lee and Shim (2019), it was utilized a Faster R-CNN to isolate potential fire areas, leveraging both spatial and temporal attributes and minimize incorrect positive identifications. Several researchers have also exploited variants of YOLO

for wildfire detection (Jiao et al., 2019; Wang et al., 2022). A combined learning strategy for spotting wildfires was proposed by merging the capabilities of two distinct models, YOLOv5 and EfficientNet (Xu et al., 2021). Overall object-detection based methods are optimized for overall object detection accuracy and do not effectively address the class imbalance issue, which is inherent in the datasets involving drone imageries.

2.3. Segmentation-based methods

Semantic segmentation-based methods can capture the fine-grained details of fire/smoke regions, even when they are relatively small or occluded. This level of granularity enables the model to capture the intricate shape and texture characteristics of fire/smoke. A few popular architectures of segmentation-based approaches include U-Net (Ronneberger et al., 2015), Deeplabv3+ (Chen et al., 2018), Mask R-CNN (He et al., 2017) and Fully Convolutional Network (FCN) (Long et al., 2015). These segmentation architectures use convolutional neural networks (CNNs) as backbone (Fraser et al., 2023). Deeplabv3+ (Chen et al., 2018), an updated version of Deeplabv3 (Chen et al., 2019), was used for segmentation tasks. Deeplabv3+ added a decoder network to the encoder of Deeplabv3 for better performance. A complete survey of well-established state-of-the-art techniques for semantic segmentation is given in Lateef and Ruichek (2019). Fusion Net was used for semantic segmentation of fire images (Choi et al., 2021). The popular segmentation framework Deeplabv3+ was used along with a generative adversarial network to delineate areas that contain smoke and anticipate subsequent smoke patterns across frames (Cheng et al., 2019).

Deeplabv3+ model was used for semantic segmentation of fire pixels using the Corsican dataset (Harkat et al., 2020). A combined infra-red and RGB images and applied both rule-based methods and U-Net framework was introduced to segment fire/smoke pixels (Kacker et al., 2023). These research works employed segmentation frameworks with CNN backbones. However, segmentation techniques using CNN-based backbone network struggle to capture long-range dependencies between objects in an image.

Vision transformers (ViTs) (Dosovitskiy et al., 2020) are used as a replacement to CNNs for expansive image recognition tasks by exploiting its self-attention mechanisms. Mobile vision transformer (MobileViT) (Mehta and Rastegari, 2021) is an architecture that uses ViT with CNN to overcome the drawbacks of ViTs. MobileViT with Deeplab (Chen et al., 2017) were integrated for segmentation tasks and showed promising results on PASCAL VOC 2012 dataset. In term of forest fire segmentation, two vision-based Transformers, TransUNet and MedT, were proposed using the Corsican dataset (Ghali et al., 2021) achieving high competitive results by increasing the model complexity.

2.4. Challenges and contributions

In view of the above, the following challenges are identified in current fire detection systems using drones.

1. Colour-based approaches may produce high number of false positives due to an inherent cognitive bias problem.
2. High accurate deep learning models for image segmentation are computational intensive and cannot be integrated into drones.
3. Classification-based methods do not pinpoint the fire in an image or bring out its characteristics (e.g., spread, expanse), and are susceptible to high false positives.
4. Object detection-based methods are computational expensive and are susceptible for class imbalance datasets. In addition, they fail to distinguish overlapping/occluded objects and may not capture accurately the shape of fire or smoke.
5. Segmentation-based methods using CNN struggle to capture long-range dependencies between objects in an image.

Here, we are interested in two specific challenges in current fire and smoke detection algorithms

1. Expert rules have cognitive bias which leads to a large number of false positives depending on illumination conditions or objects with similar colour spectrum.
2. CNN based segmentation methods are computationally intensive and hence, cannot be applied in memory-restricted devices such as drones. In addition, they cannot capture long-range dependencies between objects in imagery data (Harkat et al., 2020).

To deal with this problem, this paper introduces two lightweight deep learning architectures for wildfire detection using drones. This is done by using a modified version of Deeplabv3+ that maintains its detection performance with less computational power and the need of expert colour-based rules. Hence, we improved the semantic segmentation architecture of Deeplabv3+ by using lightweight drone-compatible methods. This is done by improving the encoder of deeplabv3+ with two different network architectures capable to capture long-range dependencies and grasp multi-scale contextual nuances in imagery data with less number of parameters. This implies that the final model architecture is capable to accelerate the inference speed which is coherent with the real-time monitoring requirements of wildfire using drones.

The proposed models are compared with the state-of-the-art model Deeplabv3+ with the CNN-based ResNet-50 as encoder backbone, which has demonstrated to have the high accuracy detection results (Harkat et al., 2020). Colour-based approaches are not compared to ensure a fair comparison and avoid cognitive bias.

The main contributions of the proposed architectures based on the aforementioned gaps are: (1) They do not require expert colour rules, (2) They overcome the limitation of CNN-based deep learning methods by providing detailed spatial context of fire or smoke, (3) They handle the class imbalance, detect overlapping and occluded objects issues by pinpointing the location of fire or smoke accurately, and (4) They are computationally efficient and they are able to harness long-range dependencies.

3. Methodology

The proposed methodology is inspired in the Deeplabv3+ for semantic segmentation. Here, the DCNN backbone in the encoder of Deeplabv3+ framework (see Fig. 1) are heavy and computationally intensive (Mehta and Rastegari, 2021; Howard et al., 2019), making them unsuitable for employment on drones. Hence, the key idea of the proposed innovation is to improve the capabilities of the encoder in Deeplabv3+ to capture long-range dependencies by either 1) integrating a self-attention-centric vision transformer (MobileViT) or 2) using a light version of Mobilenetv3. By enhancing the encoder capabilities, the proposed networks can grasp multi-scale contextual nuances in an image more proficiently, thereby amplifying the proficiency of the network in discerning both large-scale patterns and intricate spatial specifics. Additionally, the proposed architectures have less number of parameters and faster inference capabilities, making them highly suitable for real-time wildfire monitoring tasks using drones.

3.1. Deeplabv3+

The proposed approach uses as baseline the deeplabv3+ for image segmentation. Deeplabv3+ has an encoder-decoder structure whose diagram is observed in Fig. 2.

The encoder is composed of two main segments: the DCNN backbone model and the ASPP (Atrous Spatial Pyramid Pooling) block. The ASPP block allows the network to successfully harness multi-scale context by enlarging the field of view of the filter (Chen et al., 2017). Within this framework, the encoder is responsible for harnessing intricate contextual nuances in the input image by scrutinizing its

characteristics at multiple scales. For this, it uses a variety of filters and pooling techniques across diverse resolutions. Consequently, the decoder recovers the spatial information gradually to capture sharp boundaries in input images. Here, the DCNN backbone model plays a pivotal role in the success of Deeplabv3+ for image segmentation since it is responsible to extract high-dimensional features from a given image that feed both the ASPP and decoder blocks for accurate detection of fire. Therefore, we are interested in preserving the good detection properties of Deeplabv3+ by improving its long-range detection capabilities between objects in the image, whilst reducing the complexity of the DCNN backbone model.

3.2. Proposed MobileViT-Deeplabv3+ architecture

The first approach that we adopt is modifying the DCNN backbone block of Fig. 2 with an improved architecture based on the MobileViT model proposed in Mehta and Rastegari (2021). The key idea of MobileViT is the incorporation of inverted residual blocks of mobilenetv2 (Sandler et al., 2018) with a novel “MobileViT block”. This integration allowed the MobileViT block to exhibit properties of both CNNs and ViTs, leveraging the spatial inductive biases of CNNs and the input-adaptive weighting and global processing capabilities of ViTs.

Fig. 3 depicts the elements of each MobileViT block exhibited in the MobileViT architecture (for more details refer to Dosovitskiy et al. (2020) and Mehta and Rastegari (2021)). X and Y represent input and output tensors, respectively, with height H , width W and C denotes the number of channels. $n \times n$ Conv represents standard $n \times n$ convolution operation. The principle of operation of the MobileViT Block is divided into the following four sections:

- **Local Representation Operation:** MobileViT block applies a standard $n \times n$ convolution layer followed by a pointwise (or 1×1) convolutional layer to an input tensor X of dimensions $H \times W \times C$. This gives a tensor X_i of dimension $H \times W \times d$, where d is a higher-dimensional space, such that $d > C$.
- **Unfolding Operation:** The tensor X_i of dimension $H \times W \times d$ is then decomposed and flattened into N disjoint patches, generating a tensor X_U of dimension $P \times N \times d$, where P is the breakdown of patches having dimension $w \times h$ (width w and height h of each patch). Unfolding enables MobileViT to learn global representations with spatial inductive bias. The unfolded tensor X_U is then passed through the transformer of L blocks. Transformer encodes inter-patch relationship to each patch.
- **Folding Operation:** Given that MobileViT retains the sequence of each individual patch and the inherent spatial arrangement of pixels inside them, the tensor X_U can be seamlessly transformed back, resulting in the tensor X_i of dimension $H \times W \times d$. This characteristic ensures that the spatial coherence and structure are preserved even after processing through the transformer layers, facilitating accurate spatial feature representations in the subsequent stages of the network.
- **Fusion Operation:** A pointwise convolution is applied to this tensor X_i to project it to a low C -dimensional space. This gives a tensor of dimension $H \times W \times C$ which is concatenated with input tensor X to learn low-level features. We now obtain tensor $H \times W \times 2C$. Another $n \times n$ convolution is applied to fuse the concatenated tensors and we get the output tensor Y with same dimensions as the input tensor X .

The overall architecture of the proposed MobileViT-Deeplabv3+ is depicted in Fig. 4. In this paper, we have two variants of the MobileViT architecture: MobileViT_S and MobileViT_XS. These variants differ in accordance with the number of channel and projection dimensions. These parameters dictate the depth of MobileViT model, and the dimension of the output feature maps at different spatial resolutions, enabling the model to capture global dependencies and process the

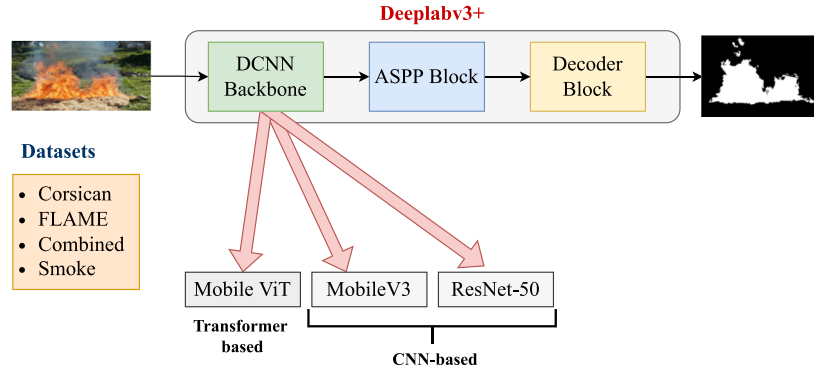


Fig. 1. System architecture overview.

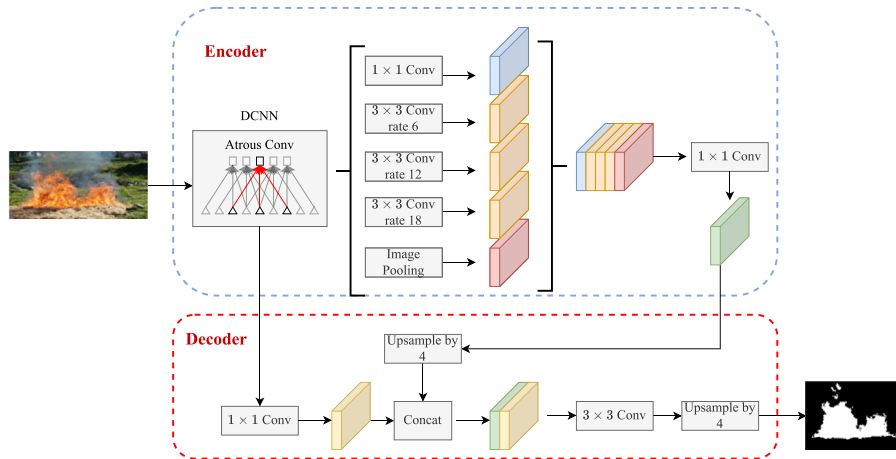


Fig. 2. Encoder-Decoder architecture of Deeplabv3+ adapted from Chen et al. (2018).

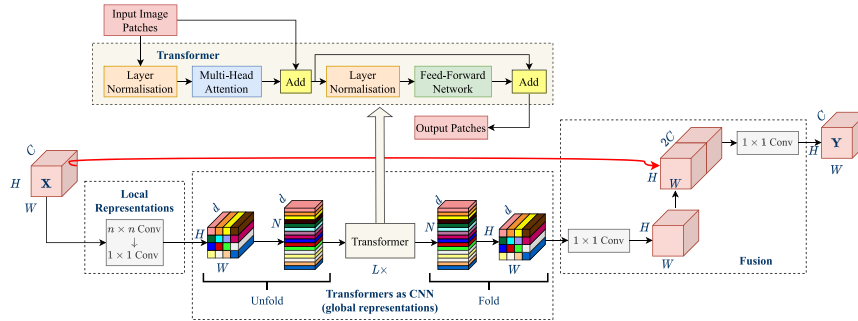


Fig. 3. Elements of the Mobile ViT block.

Source: Adapted from Mehta and Rastegari (2021).

input data effectively. Here, MobileViT_S has [144, 192, 240] projection dimensions and [16, 32, 64, 64, 96, 128, 160, 640] channels. On the other hand, MobileViT_XS has [96, 120, 144] projection dimensions and the number of channels are [16, 32, 48, 48, 64, 80, 96, 384]. This clearly shows a reduction of the number of parameters of the network required for drone applications.

3.3. Proposed Mobilenetv3-Deeplabv3+ architecture

The second approach consists in maintaining the CNN topology of the DCNN but using a lightweight deep model version that can be implemented in drones. Mobilenetv3 (Howard et al., 2019) has been used with Deeplabv3+ for wildfire segmentation (Li et al., 2022) by using two low-level features to the decoder module which do not exploit the complete spatial information of the image. Therefore, the main novelty

of the proposed architecture is to solve the spatial weakness observed in Li et al. (2022) by removing the low-level features in the decoder and incorporate an ASS block to maintain spatial resolution while being able to recognize larger-scale features. The proposed approach is shown in Fig. 5. Notice that the ASPP block of deeplabv3+ is slightly modified with depthwise separable 2D convolutions (SepConv) for better spatial recognition on the image.

The architecture of the Mobilenetv3 bottlenecks (based on inverted residual blocks, see Fig. 4) are summarized in Table 1.

3.4. Architectures analysis

Both the proposed MobileViT-Deeplabv3+ and Mobilenetv3-Deeplabv3+ architectures show high reduction of the number of parameters of the overall network. Here both MobileViT and Mobilenetv3

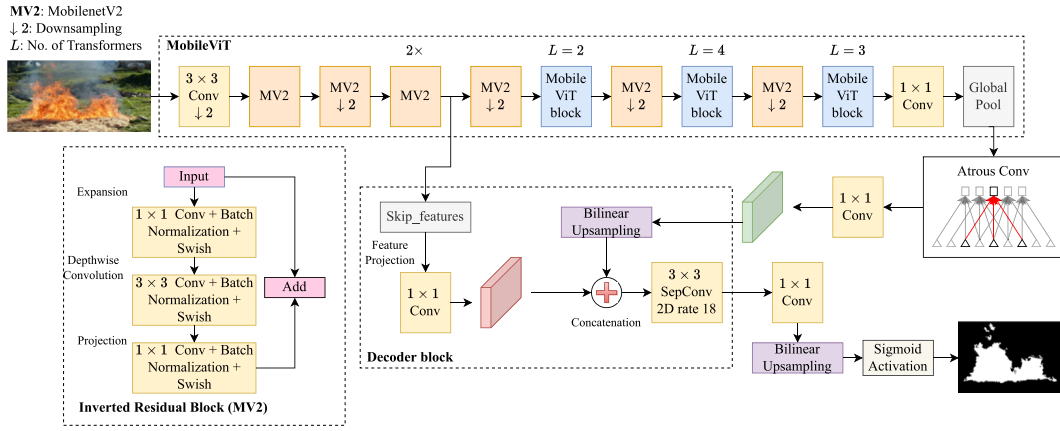


Fig. 4. Proposed MobileViT-DeeplabV3+ architecture.

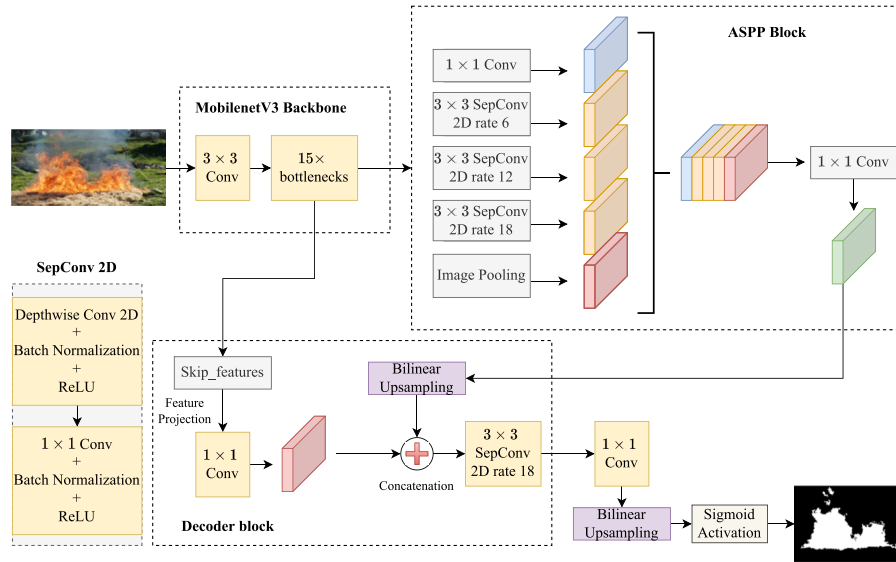


Fig. 5. Overall architecture of Mobilenetv3-DeeplabV3+.

Table 1
Mobilenetv3 bottlenecks architecture.

Input	Operator	Kernel size	Expansion size	Filters	Squeeze & Excite modules	Activation	Stride
$512 \times 512 \times 3$	Conv2D	3×3	–	16	No	HS	2
$256 \times 256 \times 16$	BottleNeck	3×3	16	16	–	ReLU	1
$256 \times 256 \times 16$	BottleNeck	3×3	64	24	No	ReLU	2
$128 \times 128 \times 24$	BottleNeck	3×3	72	24	No	ReLU	1
$128 \times 128 \times 24$	BottleNeck	5×5	72	40	Yes	ReLU	2
$64 \times 64 \times 40$	BottleNeck	5×5	120	40	Yes	ReLU	1
$64 \times 64 \times 40$	BottleNeck	5×5	120	40	Yes	ReLU	1
$64 \times 64 \times 40$	BottleNeck	3×3	240	80	No	HS	2
$32 \times 32 \times 80$	BottleNeck	3×3	200	80	No	HS	1
$32 \times 32 \times 80$	BottleNeck	3×3	184	80	No	HS	1
$32 \times 32 \times 80$	BottleNeck	3×3	184	80	No	HS	1
$32 \times 32 \times 80$	BottleNeck	3×3	480	112	Yes	HS	1
$32 \times 32 \times 112$	BottleNeck	3×3	672	112	Yes	HS	1
$32 \times 32 \times 160$	BottleNeck	5×5	672	160	Yes	HS	1
$32 \times 32 \times 160$	BottleNeck	5×5	960	160	Yes	HS	1
$32 \times 32 \times 160$	BottleNeck	5×5	960	160	Yes	HS	1

enhance the detection properties of the DCNN block by discerning both large-scale attributes and intricate spatial details, simultaneously. This contributes in the effectiveness of the decoder block in the reconnaissance of spatial fire and smoke features and also capturing long-range dependencies. The use of inverted residual blocks allow to maintain specific features required to grasp global representations and obtain accurate image segmentation results.

4. Datasets

Obtaining a dataset that is sufficiently large and diverse is a huge challenge in wildfire and smoke detection tasks. Most of the datasets are fire-centric and ignore smoke. Furthermore, there is no publicly available dataset that has both fire and smoke images together in a single frame, taken from drones in the context of forests. Hence, we resorted

Table 2
Pixel distribution across all the datasets.

Dataset	Total images	Fire/Smoke pixels	Background pixels	Number of fire pixels	Number of background pixels
Corsican	1122	17.09%	82.91%	256.43M	1243.9M
FLAME	2003	0.58%	99.42%	97.12M	16516.55M
Combined	3125	1.95%	98.05%	353.56M	17760.45M
Smoke	2848	12.44%	87.56%	108.91M	766.41M

to binary semantic segmentation of fire and smoke, independently. For fire segmentation task, we have used three datasets

1. The Corsican dataset (Toulouse et al., 2017). Consisting of 1135 RGB and 640 infrared image of fire across different environmental settings. Additionally, it contains segmentation ground truths for fire. Here, only 1122 RGB images and corresponding masks were selected for this study. These images were chosen since do not contain objects that can produce bias issues. Fire in these images are large and occupy a relative proportion of the image.
2. The FLAME (Fire Luminosity Airborne-based Machine Learning Evaluation) dataset (Shamsoshoara et al., 2021). Consists in 10 repositories containing images and videos for different fire-related tasks. For this study, the ninth repository is used with 2003 images of fire and corresponding masks. Here, we select the ninth repository because it includes images for segmentation tasks. Fire in these images are small and occupy a low proportion of the image.
3. A combined dataset. The Corsican and FLAME datasets are combined to obtain a better representation of different environmental settings with fire images taken from different angles. The obtained dataset consists of 3125 images with their respective masks.

Notice that the Corsican and FLAME datasets focus on fires with different proportion of the image. This means that a network trained with Corsican dataset cannot easily detect small fires. Conversely, a network trained purely with FLAME dataset struggles to accurately detect large fires. To solve this issue, we create a combined dataset that enables better generalization.

For segmenting smoke, a custom smoke dataset was created. In this case, two datasets are used: SMOKE5K and AI-For-Mankind. On the one hand, the SMOKE5K dataset contains 1400 real and 4000 synthetic images. These synthetic data are discarded since they do not comply with the requirements of smoke detection in forests using imagery data captured by a drone (e.g., they are images from indoor places). On the other hand, manual annotation (using LabelMe tool) is conducted on the images of the AI-For-Mankind to obtain the masks of smoke. Here, we ensure that the images cover diverse smoke scenarios. Images with varying levels of smoke density, ranging from light haze to thick plumes were considered for better generalization of the final models. Images with diverse climatic and lightning conditions were also included. The final dataset consists of 2848 images. Fig. 6 shows some image samples of each dataset.

Class imbalance in drone-captured datasets is a prevalent issue in image segmentation, especially when the target- smoke or fire, is smaller or sparser compared to the background. Such imbalance can bias the model towards predicting the dominating class more often, undermining its performance on the under represented class. Here, all the datasets used in this work exhibit significant pixel-level imbalance with the majority class (background) dominating over the minority class (fire/smoke). Table 2 summarizes the pixel-distribution across all the datasets.

The results indicate that our model will predominantly encounter background pixels during training. Consequently, its proficiency in detecting and learning from the minority class, which is the fire/smoke

pixels, might be hindered. This could lead to suboptimal performance in recognizing fire regions in new, unseen images. The images and masks are scaled into images of $512 \times 512 \times 3$ and $512 \times 512 \times 1$, respectively. We observe that this size produces better results in the segmentation tasks.

4.1. Regularization and augmentation

L_2 regularization has been used in the convolutional layers of the ASPP and the Decoder blocks in Deeplabv3+. This encourages the model to learn smaller weights, which shall help to prevent the overfitting problem. To bolster the representation of the minority class, we employed data augmentation techniques such as: flipping, rotations and translations to be consistent with different imagery data captured by drones. These approaches can mitigate the challenges introduced by class imbalance and facilitate the model's accuracy in distinguishing and learning from the minority class. By providing a more balanced representation of the classes during training, the model can identify the fire regions efficiently.

4.2. Loss functions

We have considered five commonly utilized loss functions for image segmentation tasks. The loss functions considered in this work are: Dice, Tversky, Focal Tversky, Binary Cross-Entropy, and a combination of dice loss and binary cross-entropy loss (Combo loss) defined as

$$\text{Combo Loss} = \alpha \cdot \text{Binary Cross Entropy Loss} + \beta \cdot \text{Dice Loss}, \quad (1)$$

where $\alpha, \beta \in [0, 1]$ are user-define weights that are adjusted to balance the contribution between both losses.

5. Experiments

The following architectures were tested under the proposed datasets for fire and smoke detection: MobileViT_S-Deeplabv3+, MobileViT_XS-Deeplabv3+, Mobilenetv3-Deeplabv3+, ResNet50-Deeplabv3+.

We evaluated the proposed architectures across each dataset separately. We assessed the performance of each proposed architecture in combination with a variety of loss functions, using a comprehensive suite of metrics, including Mean Pixel Accuracy (MPA), F1 score, Intersection over Union (IoU), Recall, Precision, and Matthews Correlation Coefficient (MCC), for a robust and extensive comparative study. The architectures were implemented using TensorFlow (version 2.12.0) in Python (version 3.8.10). The models were trained using four NVIDIA A100-PCIE GPUs with memory of 40 GB each. Here, the objective of the experiments is twofold: 1) demonstrate the effectiveness of the proposed approach for fire and smoke segmentation with less computational complexity and 2) identify which model behave better given a specific loss function and metrics.

Throughout the experiments, several control variables were kept constant to ensure that the comparison between the models are fair and unambiguous. We split each dataset in the ratio of 70-15-15, i.e., 70% training, 15% validation and 15% testing. Subsequently, the training data was subjected to the following augmentations to increase the diversity and address the class imbalance- horizontal flipping, random rotation, random brightness, and random contrast. The Adam optimizer was used in the training of all the models, owing to its proven efficiency and its adaptive learning rate properties. A learning rate of $1e^{-4}$ was selected, which offers a good balance between training speed and model accuracy. Furthermore, the learning rate was kept adaptive and reduced by a factor of 0.1 if the training did not improve for 5 epochs. Additionally, to maintain the consistency of input size across all the models, a batch size of 16 was used. We chose the output stride as 16 with dilation rates 6, 12 and 18 for the ASPP block of Deeplabv3+. The

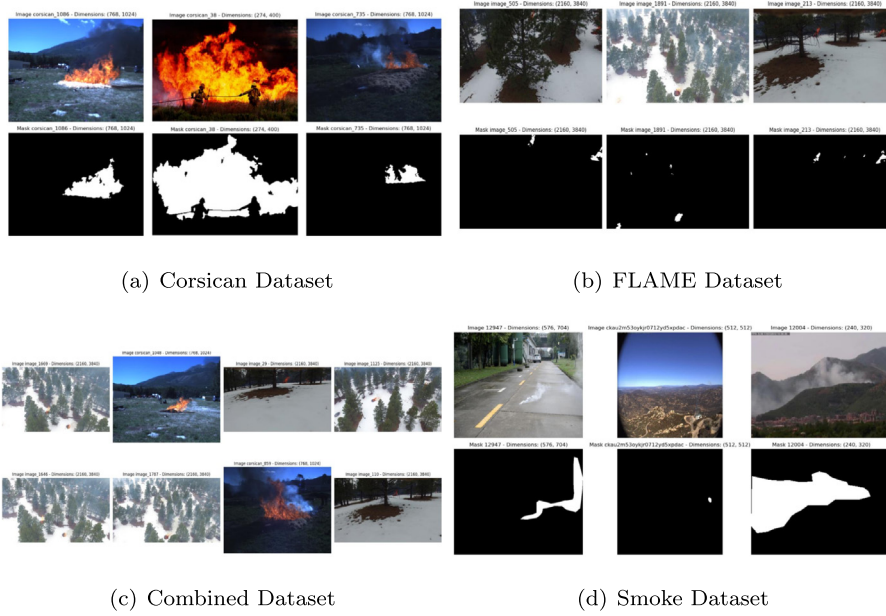


Fig. 6. Image samples of each dataset with their respective mask used in this research.

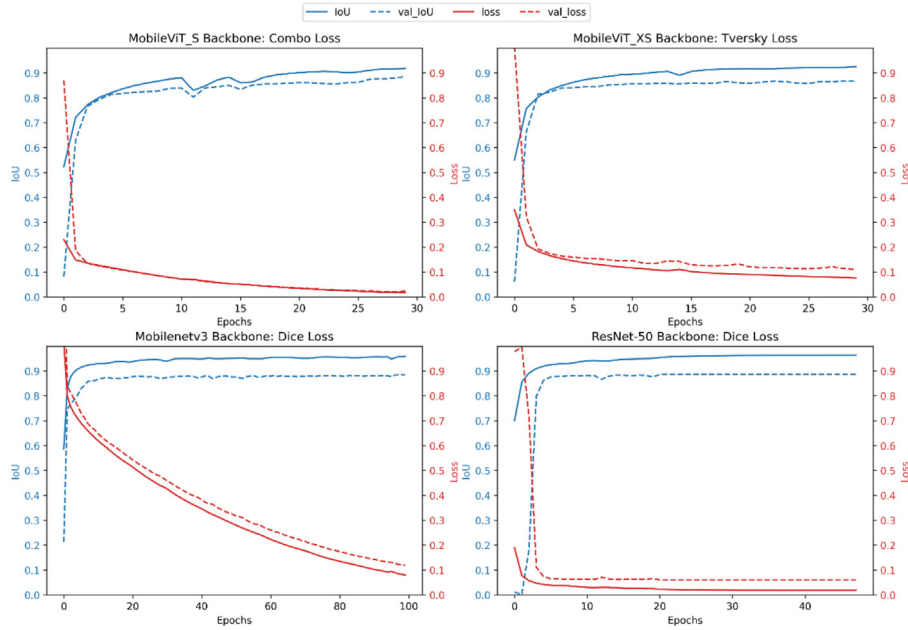


Fig. 7. IoU vs. Loss curves for the best performing loss functions on the Corsican dataset.

number of epochs were set to 30 for all models except for MobilenetV3-Deeplabv3+ which requires more epochs to converge. The number of epochs for MobilenetV3-Deeplabv3+ was set to 100.

We applied a small amount of L_2 regularization ($2e^{-5}$) to penalize the model lightly for having large weight coefficients, helping prevent overfitting while still allowing the model to learn from the data. Furthermore, for the Tversky Loss the values of α and β were kept constant at 0.3 and 0.7, respectively for all the datasets. The optimal outcomes using Focal Tversky loss were achieved by maintaining a value of $\gamma = 2.0$.

5.1. Results

Fig. 7 and Table 3 gives the best performing loss functions for each model in the Corsican dataset. It is observed that the proposed models

Table 3

Best performing loss functions on the Corsican dataset.

Backbone	Loss	MPA	F1	IoU	Recall	Precision	MCC
MobileViT_S	Combo	0.9876	0.9332	0.8773	0.9613	0.9502	0.9251
MobileViT_XS	Tversky	0.9869	0.9295	0.8722	0.9552	0.9099	0.9217
MobilenetV3	Dice	0.9879	0.9365	0.8824	0.9439	0.9323	0.9287
ResNet-50	Dice	0.9886	0.9402	0.8888	0.9472	0.9353	0.9324

do not have overfitting problem and exhibit good generalization. Whilst the dice loss in tandem with the ResNet50-based architecture gave the best results, the performances of other models are not far behind. Specifically, the MobileViT_S model exhibited the highest precision and recall values.

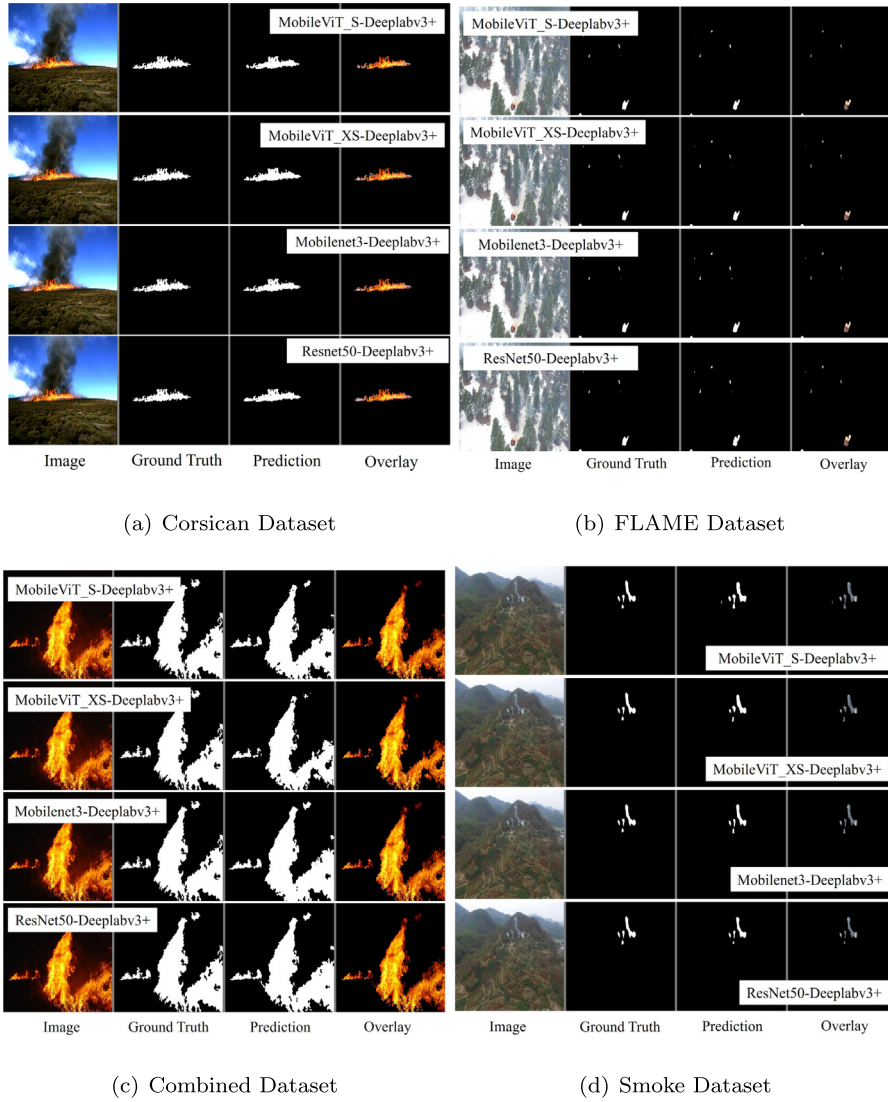


Fig. 8. Predictions on correctly detected samples across all datasets.

Fig. 8(a) shows the segmentation results under the Corsican dataset. Upon closer examination, it can be observed that all the models have performed commendably, yielding high-quality segmentation that closely matches the ground truth. Despite differences in the model architectures and the loss functions, the predicted segmentation exhibits a high degree of consistency across all the models. This observation reinforces the fact that all the proposed models are equally competent at early detection of wildfire and smoke in the Corsican dataset.

Fig. 9 and Table 4 show the best performing loss functions for each model in the FLAME dataset. Similarly to the performance of models on the Corsican dataset, the proposed models do not over-fit to the training data. It was observed that the ResNet50-based backbone architecture delivered better performance across all the metrics for the FLAME dataset. However, the highest precision was achieved by the MobileViT_XS model, showcasing its capability to limit the false positives effectively. Fig. 8(b) shows the detection results using the FLAME dataset. We observed that, in stark contrast to their performance on the Corsican dataset, all the models trained on the FLAME dataset were able to effectively detect both small fire pixels as well as the occluded fire. On further evaluation of the trained models on unrelated fire images and a subset of Corsican dataset, we observed that the models could effectively detect fire where the proportion of fire pixels in the image are very less but struggled to detect large fires. All the models were

Table 4

Best performing loss functions on the FLAME dataset.

Backbone	Loss	MPA	F1	IoU	Recall	Precision	MCC
MobileViT_S	Combo	0.9988	0.8861	0.7992	0.8765	0.8998	0.8865
MobileViT_XS	BCE	0.9988	0.8876	0.8012	0.8792	0.9015	0.8884
MobilenetV3	Dice	0.9989	0.8931	0.8114	0.9040	0.8877	0.8938
ResNet-50	Dice	0.9989	0.8954	0.8140	0.9004	0.8937	0.8957

observed to be biased towards detecting small fires (lesser number of fire pixels in a frame compared to the background).

Fig. 10 and Table 5 show the best performing loss functions for each model in the Combined dataset. Overall, the MobileViT_XS model with BCE loss function scored higher across most metrics when compared to the other architectures without overfitting problem. This demonstrates that transformer-based architectures could deliver superior results in this specific semantic segmentation task which addresses the bias problem in the Corsican and the FLAME dataset.

Fig. 8(c) shows the detection results using the combined dataset. We observed that the models effectively detected both large fires and small, occluded fires. The performance was further tested on unseen images which had yielded poor results in the Corsican and the FLAME datasets earlier. It was observed that the models trained on the Combined dataset could effectively address these challenges. By integrating

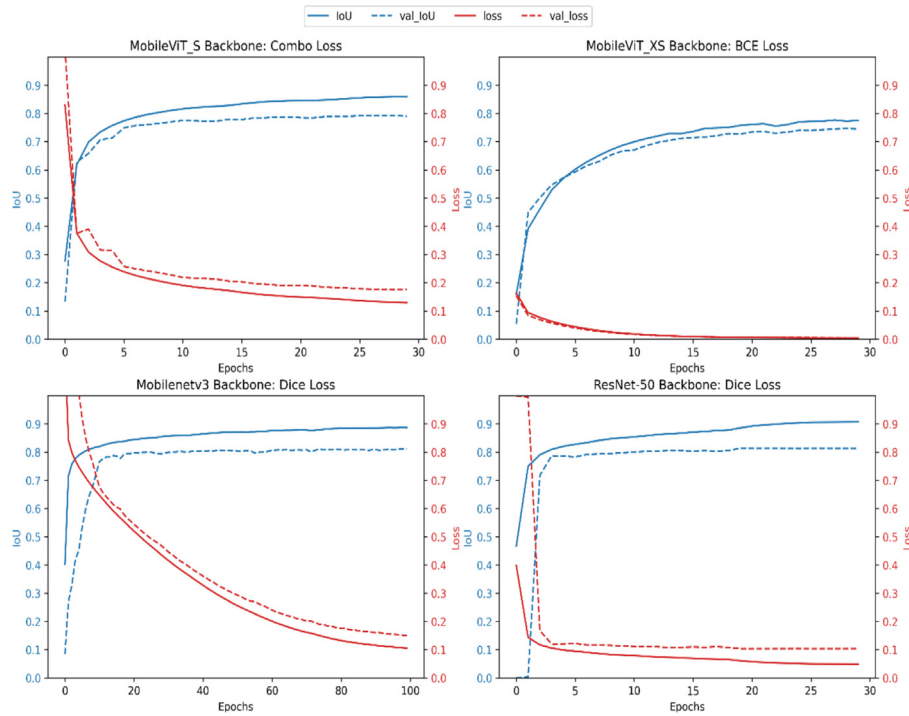


Fig. 9. IoU vs. Loss curves for the best performing loss functions on the FLAME dataset.

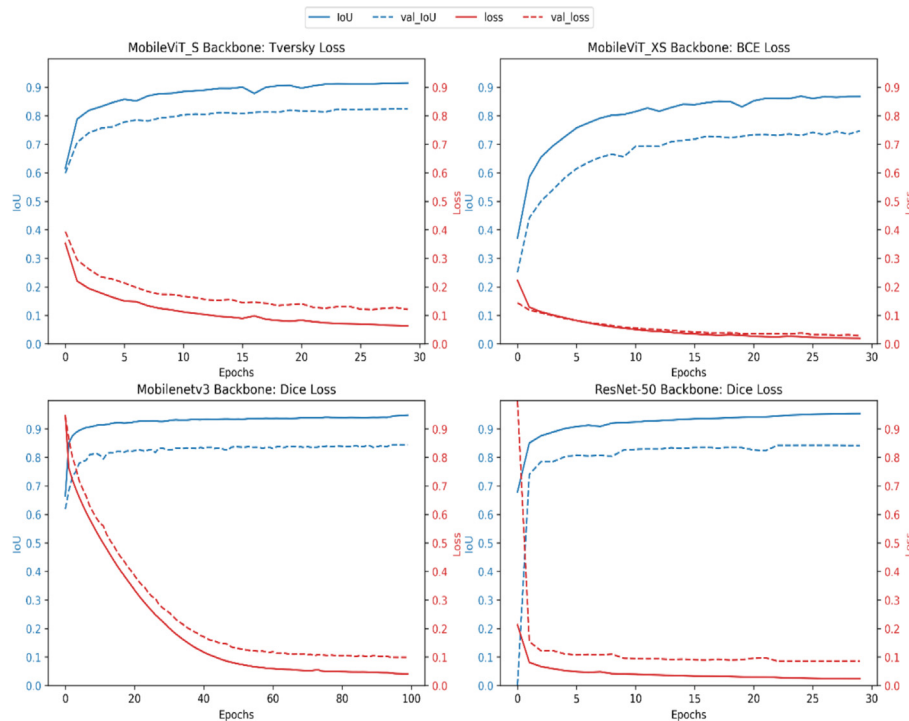


Fig. 10. IoU vs. Loss curves for the best performing loss functions on the Combined dataset.

the fire instances from both the Corsican and the FLAME datasets, the models could learn a diverse range of fire scenarios, spanning from vast, raging blazes to subtle, tiny flames. This comprehensive training pool enriched the ability of the models to discern various fire attributes, from size and intensity to colour and shape. Consequently, the Combined dataset allowed the models to generalize well across both large and small fires, as well as the occluded ones.

The performance boost on unseen images, which previously gave poor results, is a testament to this improved generalization capability. Thus, it is concluded that the performance of the models can be significantly influenced by the dataset they are trained on. The Combined dataset seems to balance the biases of the models trained exclusively on either the Corsican or the FLAME dataset. Across the board, all models and loss functions achieved high MCC values, indicating their

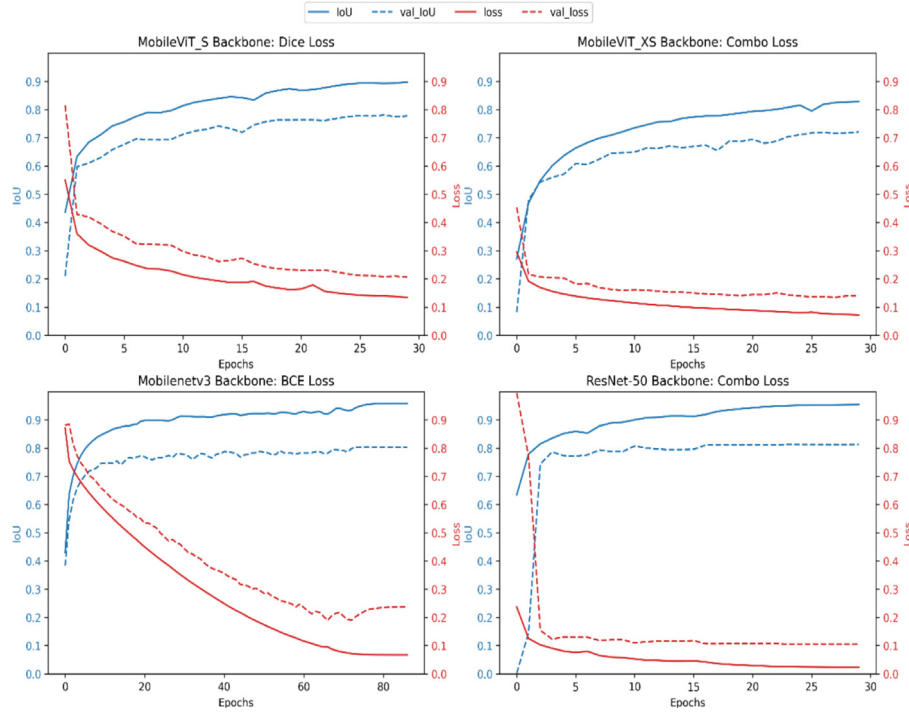


Fig. 11. IoU vs. Loss curves for the best performing loss functions on the smoke dataset.

Table 5

Best performing loss functions on the Combined dataset.

Backbone	Loss	MPA	F1	IoU	Recall	Precision	MCC
MobileViT_S	Combo	0.9932	0.8959	0.8165	0.9243	0.8741	0.8930
MobileViT_XS	BCE	0.9988	0.9286	0.8697	0.9122	0.9507	0.9207
MobilenetV3	Dice	0.9941	0.9071	0.8340	0.9006	0.9173	0.9040
ResNet-50	Dice	0.9942	0.9087	0.8371	0.9067	0.9151	0.9057

Table 6

Best performing loss functions on the Smoke dataset.

Backbone	Loss	MPA	F1	IoU	Recall	Precision	MCC
MobileViT_S	Dice	0.9718	0.8329	0.7444	0.8472	0.8458	0.8231
MobileViT_XS	Combo	0.9729	0.8289	0.7417	0.8288	0.8560	0.8195
MobilenetV3	BCE	0.9726	0.8503	0.7588	0.8283	0.9041	0.8410
ResNet-50	Combo	0.9765	0.8879	0.8091	0.9013	0.8877	0.8772

proficiency in balanced binary classifications – a trait indispensable for real-world applications.

The best performing loss functions for all the models are shown in Fig. 11 and Table 6. We observed that most models trained well on the Smoke dataset, maintaining a good balance between bias and variance. However, the performance of Mobilenetv3 exhibited a different trend. Starting from the initial stages of training, Mobilenetv3 showed a gradually decreasing trend in training loss. However, after approximately 70 epochs, we began to observe signs of overfitting. Despite maintaining a consistently low training loss, the validation loss of the model started to increase.

While all the models performed relatively well, the ResNet-50 model was able to slightly outperform the other models. Mobilenetv3 displayed notably high precision scores across all the loss functions, especially with the Combo loss, where its precision is highest. This suggests that, for this dataset and task, Mobilenetv3 is particularly adept at reducing false positives. Overall, there does not seem to be a single “best” loss function for all the models; the optimal loss function varies depending on the model in question.

Fig. 8(d) shows the detection results using the smoke dataset. We observed that the models were able to effectively segment smoke when

Table 7

Total number of parameters per model.

Architecture	Mobile-ViT_S Deeplabv3+	Mobile-ViT_XS Deeplabv3+	MobilenetV3 Deeplabv3+	Resnet50 Deeplabv3+
Parameters	6,624,689	3,064,401	3,538,065	17,869,697

the background does not have other objects with close resemblance to smoke such as fog, haze, or cloud. Furthermore, all the models could effectively detect smoke, even when they were presented with foggy backgrounds.

In our evaluation of the performance of the models on unseen smoke images, an essential observation was made that reinforces the reliability of our models. Notably, no model exhibited a bias towards detecting objects of grey colour as smoke. This is particularly significant as it indicates the models’ capability to discern smoke based on more than just colour. In real-world scenarios, this ensures that the models will not generate false alarms simply because grey-coloured objects are present in the frame. This shall enhance the practical applicability and trustworthiness of the proposed architectures.

5.2. Parameter on model efficiency

In this paper, we proposed lightweight Deeplabv3+ models using MobileViT_S, MobileViT_XS, and Mobilenetv3 backbones to effectively carry out the binary semantic segmentation task with considerably lesser number of parameters compared to the traditional ResNet-50 model. Interestingly, these models achieved performance comparable to the ResNet-50 across all the datasets and loss functions. Table 7 shows the number of parameters across all the models.

Thus, it brings an interesting observation that a transformer-based model with fewer parameters can perform similar to a traditional CNN-based model with significantly large number of parameters. The MobileViT models, for instance, leverage the Transformer architecture, which is known for its capability to harness long-range dependencies, providing a global understanding of the image that is beneficial for

Table 8
Best performing loss functions.

Backbone architecture	Datasets			
	Corsican	FLAME	Combined	Smoke
Mobile ViT_S	Combo	Combo	Tversky	Dice
Mobile ViT_XS	Tversky	BCE	BCE	Combo
Mobilenetv3	Dice	Dice	Dice	BCE
Resnet50	Dice	Dice	Dice	Combo

semantic segmentation. On the other hand, the MobilenetV3 model makes use of light CNN layers which are more computationally efficient while still maintaining a satisfactory level of representational power. This study also revealed a crucial balance between the intricacy of the model and the computational efficiency. This trade-off is particularly relevant in the context of mobile and edge computing applications, where the available computational resources are often limited. By employing lightweight architectures such as MobileViT and Mobilenetv3, we demonstrated that it is possible to develop efficient models suitable for drone applications without sacrificing performance.

5.3. Discussions and limitations

The proposed models align with the current needs of AI in engineering in the development of high accurate and lightweight models that can be used in the transportation sector. In contrast to colour-based approaches that have an inherent cognitive bias due to the use of expert rules that create many false positives, the proposed approach overcomes this issue by exploiting the capabilities of vision transformers and spatial context provided by the deeplabv3+ architecture. Additionally, the proposed methodology has a natural mechanism to reduce the number of parameters of the model, whilst maintaining competitive results.

All the datasets used in this paper were imbalanced, with fewer fire/smoke pixels compared to the background. Hence, five different loss functions were used, which effectively addressed the problem, and we did not encounter biased predictions towards background pixels during training or evaluation. Instead, our experiments revealed interesting dynamics and biases which are inherent in the characteristics of the datasets used. This introduced bias related to size and occlusion in the models during training. When the models were trained on the Corsican dataset, they demonstrated certain limitations, particularly when detecting small or obscured fires. Conversely, when trained on the FLAME dataset, the models efficiently detected both small and occluded fires, a stark contrast to their performance on the Corsican dataset. However, they struggled when faced with larger fires. The results highlight the significant impact the training dataset has on model performance. The combined dataset effectively counterbalanced the biases observed in the models trained solely on either the Corsican or the FLAME datasets, leading to more reliable and balanced binary classifications. This is particularly crucial for real-world applications where accuracy and reliability can be of paramount importance.

It was observed that different loss functions performed differently with different architectures and on different datasets. Each model seemed to have its unique ‘preference’ for a loss function that worked best as illustrated in Table 8. This could be due to how different architectures handle the gradients during the backpropagation.

Each proposed architecture has a unique layout and mechanism of learning features, which can significantly influence the rate of convergence. For all datasets, we observed that the Mobilenetv3 model took longer time to train compared to MobileViT-based architectures. A possible reason could be that the MobileViT models employ a transformer-based architecture, which excels in capturing long-range dependencies across inputs by using attention mechanisms. This allows it to learn salient features more efficiently, contributing to its faster convergence. On the other hand, the Mobilenetv3 model, which is a CNN, learns

spatial hierarchies and hence need more epochs to learn complex representations.

Overall, these findings indicate that our proposed models, with their reduced parameter count, can deliver competitive results, thus making them attractive for lightweight applications, such as drone-based wildfire detection. The findings highlight the importance of considering the interplay between the choice of loss function and the model architecture when striving to optimize performance in semantic segmentation tasks. The discussion thus demonstrates that the relevance and importance of our research extend beyond academia, providing meaningful contributions to the wider field and society.

6. Conclusions

In this paper, two lightweight DL models for wildfire and smoke segmentation using drones are proposed. The approach uses as baseline the Deeplabv3+ architecture for image segmentation. The computational and long-range dependencies weakness are addressed by updating the DCNN backbone architecture with either transformer-based or CNN-based architectures that gives detailed spatial context and accurate localization of fire/smoke, whilst it reduces the computational complexity.

Despite having significantly fewer parameters, these models exhibited performance comparable to the traditional ResNet50-Deeplabv3+ architecture. Specifically, MobileViT_S had roughly one-third of the parameters of the ResNet50 model, while MobileViT_XS and Mobilenetv3 models had about one-sixth of the parameters. From the obtained results, it is clear that the proposed architectures will benefit search and rescue applications by equipping drones with lightweight models that effectively and efficiently detect fire for its early mitigation. This in consequence will reduce the negative impact of fire on national reserves and ecosystems.

The main challenge of the proposed models is the collection of appropriate data that model diverse scenarios including: large and small fires, different weather conditions, and without objects that are not relevant to our case study (e.g., indoor images of smoke or artificial smoke). This is topic of our interest that will highly contribute to the research community in this important area.

Future work will address the extension of the proposed models to perform multi-class segmentation tasks. Instead of merely identifying the presence of fire or smoke, the models could be trained to detect both smoke and fire, simultaneously within a single frame. This could potentially enhance the responsiveness and utility of these models in real-world scenarios, where early detection of both elements is critical. The integration of the fire physics as physics informed constraint for weights manifold of the network will also be investigated as further work to enhance the detection capabilities of the proposed methodology and add a new predictive dimension for early mitigation of fire.

CRedit authorship contribution statement

Abhinav Kumar: Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Adolfo Perrusquía:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Saba Al-Rubaye:** Supervision, Resources, Project administration, Conceptualization. **Weisi Guo:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Abdusalomov, A., Baratov, N., Kutlimuratov, A., Whangbo, T.K., 2021. An improvement of the fire detection and classification method using YOLOv3 for surveillance systems. *Sensors* 21 (19), 6519.
- Abdusalomov, A.B., Islam, B.M.S., Nasimov, R., Mukhiddinov, M., Whangbo, T.K., 2023. An improved forest fire detection method based on the detectron2 model and a deep learning approach. *Sensors* 23 (3), 1512.
- Akhloufi, M.A., Couturier, A., Castro, N.A., 2021. Unmanned aerial vehicles for wildland fires: Sensing, perception, cooperation and assistance. *Drones* 5 (1), 15.
- Celik, T., Demirel, H., 2009. Fire detection in video sequences using a generic color model. *Fire Saf. J.* 44 (2), 147–158.
- Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2019. Rethinking atrous convolution for semantic image segmentation. *arXiv 2017, arXiv preprint arXiv:1706.05587* 2.
- Chen, T.-H., Wu, P.-H., Chiou, Y.-C., 2004. An early fire-detection method based on image processing. In: 2004 International Conference on Image Processing, 2004. ICIP'04, Vol. 3, IEEE, pp. 1707–1710.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 801–818.
- Cheng, S., Ma, J., Zhang, S., 2019. Smoke detection and trend prediction method based on Deeplabv3+ and generative adversarial network. *J. Electron. Imaging* 28 (3), 033006.
- Choi, H.-S., Jeon, M., Song, K., Kang, M., 2021. Semantic fire segmentation model based on convolutional neural network for outdoor image. *Fire Technol.* 1–15.
- Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1251–1258.
- Chung, Y.-L., Chung, H.-Y., Chou, C.-W., 2019. Efficient flame recognition method based on a deep convolutional neural network and image processing. In: 2019 IEEE 8th Global Conference on Consumer Electronics. GCCE, IEEE, pp. 573–574.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- El Debeiki, M., Al-Rubaye, S., Perrusquía, A., Conrad, C., Flores-Campos, J.A., 2024. An advanced path planning and UAV relay system: Enhancing connectivity in rural environments. *Future Internet* 16 (3), 89.
- Fraser, B., Perrusquía, A., Panagiotakopoulos, D., Guo, W., 2023. Hybrid deep neural networks for drone high level intent classification using non-cooperative radar data. In: 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering. ICECCME, IEEE, pp. 1–6.
- Ghali, R., Akhloufi, M.A., Jmal, M., Soudene Mseddi, W., Attia, R., 2021. Wildfire segmentation using deep vision transformers. *Remote Sens.* 13 (17), 3527.
- Gruffeille, C., Perrusquía, A., Tsourdos, A., Guo, W., 2024. Disaster area coverage optimisation using reinforcement learning. In: 2024 International Conference on Unmanned Aircraft Systems. ICUAS.
- Harkat, H., Nascimento, J., Bernardino, A., 2020. Fire segmentation using a DeepLabv3+ architecture. In: Image and Signal Processing for Remote Sensing XXVI. Vol. 11533, SPIE, pp. 134–145.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.
- Hosni Mahmoud, H.A., Alharbi, A.H., Alghamdi, N.S., 2022. Time-efficient fire detection convolutional neural network coupled with transfer learning. *Intell. Autom. Soft Comput.* 31 (3).
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., 2019. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1314–1324.
- Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., Liu, D., 2019. A deep learning based forest fire detection approach using UAV and YOLOv3. In: 2019 1st International Conference on Industrial Artificial Intelligence. IAI, IEEE, pp. 1–5.
- Kacker, T., Perrusquía, A., Guo, W., 2023. Multi-spectral fusion using generative adversarial networks for UAV detection of wild fires. In: 2023 International Conference on Artificial Intelligence in Information and Communication. ICAIIC, IEEE, pp. 182–187.
- Kim, B., Lee, J., 2019. A video-based fire detection using deep learning models. *Appl. Sci.* 9 (14), 2862.
- Lakhan, A., Elhoseny, M., Mohammed, M.A., Jaber, M.M., et al., 2022. SFDWA: secure and fault-tolerant aware delay optimal workload assignment schemes in edge computing for internet of drone things applications. *Wirel. Commun. Mob. Comput.* 2022.
- Lateef, F., Ruichek, Y., 2019. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* 338, 321–348.
- Lee, Y., Shim, J., 2019. False positive decremented research for fire and smoke detection in surveillance camera using spatial and temporal features based on deep learning. *Electronics* 8 (10), 1167.
- Li, M., Zhang, Y., Mu, L., Xin, J., Yu, Z., Jiao, S., Liu, H., Xie, G., Yingmin, Y., 2022. A real-time fire segmentation method based on a deep learning approach. *IFAC-PapersOnLine* 55 (6), 145–150.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, pp. 21–37.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.
- Mehta, S., Rastegari, M., 2021. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.
- Pausas, J.G., Keeley, J.E., 2021. Wildfires and global change. *Front. Ecol. Environ.* 19 (7), 387–395.
- Perrusquía, A., Guo, W., Fraser, B., Wei, Z., 2024. Uncovering drone intentions using control physics informed machine learning. *Commun. Eng.* 3 (1), 36.
- Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7263–7271.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4510–4520.
- Sathishkumar, V.E., Cho, J., Subramanian, M., Naren, O.S., 2023. Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecol.* 19 (1), 1–17.
- Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P.Z., Blasch, E., 2021. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Comput. Netw.* 193, 108001.
- Toulouse, T., Rossi, L., Campana, A., Celik, T., Akhloufi, M.A., 2017. Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Saf. J.* 92, 188–194.
- Wang, W., Huang, Q., Liu, H., Jia, Y., Chen, Q., 2022. Forest fire detection method based on deep learning. In: 2022 International Conference on Cyber-Physical Social Intelligence. ICCSI, IEEE, pp. 23–28.
- Wu, S., Zhang, L., 2018. Using popular object detection methods for real time forest fire detection. In: 2018 11th International Symposium on Computational Intelligence and Design. ISCID, Vol. 1, IEEE, pp. 280–284.
- Xu, R., Lin, H., Lu, K., Cao, L., Liu, Y., 2021. A forest fire detection system based on ensemble learning. *Forests* 12 (2), 217.
- Yandouzi, M., Grari, M., Idrissi, I., Boukabous, M., Moussaoui, O., Azizi, M., Ghomid, K., Elmiad, A.K., 2022a. Forest fires detection using deep transfer learning. *Int. J. Adv. Comput. Sci. Appl.* 13 (8).
- Yandouzi, M., Grari, M., Idrissi, I., Moussaoui, O., Azizi, M., Ghomid, K., Elmiad, A.K., 2022b. Review on forest fires detection and prediction using deep learning and drones. *J. Theor. Appl. Inf. Technol.* 100 (12), 4565–4576.
- Yao, H., Qin, R., Chen, X., 2019. Unmanned aerial vehicle for remote sensing applications—A review. *Remote Sens.* 11 (12), 1443.
- Zunair, H., Hamza, A.B., 2021. Sharp U-net: Depthwise convolutional network for biomedical image segmentation. *Comput. Biol. Med.* 136, 104699.



DSPACE

<https://dspace.org/>

Wildfire and smoke early detection for drone applications: a light-weight deep learning approach

Kumar, Abhinav; Perrusquía, Adolfo; Al-Rubaye, Saba; Guo, Weisi

2024-10-01

Elsevier

<https://doi.org/10.1016/j.engappai.2024.108977>

Kumar A, Perrusquía A, Al-Rubaye S, Guo W. (2024) Wildfire and smoke early detection for drone applications: a light-weight deep learning approach. Engineering Applications of Artificial Intelligence, Volume 136, Part B, October 2024, Article number 108977