

Article

# Suburban Forest Fire Risk Assessment and Forest Surveillance Using 360-Degree Cameras and a Multiscale Deformable Transformer

Panagiotis Barmpoutis <sup>1,2,\*</sup>, Aristeidis Kastridis <sup>3</sup> , Tania Stathaki <sup>1</sup>, Jing Yuan <sup>1</sup>, Mengjie Shi <sup>4</sup>   
and Nikos Grammalidis <sup>2</sup> 

<sup>1</sup> Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, London SW7 2AZ, UK

<sup>2</sup> Information Technologies Institute, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece

<sup>3</sup> Laboratory of Mountainous Water Management and Control, School of Forestry and Natural Environment, Aristotle University of Thessaloniki, 54636 Thessaloniki, Greece

<sup>4</sup> Department of Biomedical Engineering & Imaging Sciences, King's College London, London WC2R 2LS, UK

\* Correspondence: p.barmpoutis@imperial.ac.uk or panbar@iti.gr

**Abstract:** In the current context of climate change and demographic expansion, one of the phenomena that humanity faces are the suburban wildfires. To prevent the occurrence of suburban forest fires, fire risk assessment and early fire detection approaches need to be applied. Forest fire risk mapping depends on various factors and contributes to the identification and monitoring of vulnerable zones where risk factors are most severe. Therefore, watchtowers, sensors, and base stations of autonomous unmanned aerial vehicles need to be placed carefully in order to ensure adequate visibility or battery autonomy. In this study, fire risk assessment of an urban forest was performed and the recently introduced 360-degree data were used for early fire detection. Furthermore, a single-step approach that integrates a multiscale vision transformer was introduced for accurate fire detection. The study area includes the suburban pine forest of Thessaloniki city (Greece) named Seich Sou, which is prone to wildfires. For the evaluation of the performance of the proposed workflow, real and synthetic 360-degree images were used. Experimental results demonstrate the great potential of the proposed system, which achieved an *F-score* for real fire event detection rate equal to 91.6%. This indicates that the proposed method could significantly contribute to the monitoring, protection, and early fire detection of the suburban forest of Thessaloniki.

**Keywords:** suburban forest fire; forest fire risk assessment; early fire detection; deep learning; deformable transformer; 360-degree cameras



**Citation:** Barmpoutis, P.; Kastridis, A.; Stathaki, T.; Yuan, J.; Shi, M.; Grammalidis, N. Suburban Forest Fire Risk Assessment and Forest Surveillance Using 360-Degree Cameras and a Multiscale Deformable Transformer. *Remote Sens.* **2023**, *15*, 1995. <https://doi.org/10.3390/rs15081995>

Academic Editor: Carmen Quintano

Received: 13 February 2023

Revised: 29 March 2023

Accepted: 2 April 2023

Published: 10 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Wildfires pose serious hazards to ecological systems and human safety. To avoid the occurrence of fire accidents, fire risk assessment, fire potential mapping, and fire surveillance systems aim to manage forests and to prevent or detect forest fires at the initial stage. Wildfire risk assessment could be defined as a combination of fire likelihood, intensity, and effects [1]. Thus, risk assessment contributes to the identification of optimal locations for the installation of fire detection sensors. Fire detection methods based on physical sensors or vision-based algorithms have been widely studied. In the past, most of the traditional detection systems were based on physical sensors. However, in the recent few decades, various kinds of vision-based security and surveillance systems have been developed. Combined with computer vision and image-processing-based methods, they can achieve accurate early fire detection, protecting human life and enhancing environmental security [2].

Various methods have been applied for wildfire risk assessment [3], which aims to minimize threats to life, property, and natural resources. The risk assessment includes

empirical or statistical studies of ignition and large-fire patterns [4] and simulation modelling where probabilities are estimated using Monte Carlo sampling of weather data [5]. In the context of fire risk assessment, both empirical and simulation studies play an important role in understanding the factors that contribute to the severity and spread of fires. Empirical studies involve collecting data through observation and analysis of real-world fire events and their impact on human and ecological systems. This type of research is crucial for gaining a deeper understanding of the various factors that contribute to the spatial planning and development of policies aimed at increasing preparedness for large fires. Such studies provide valuable insights into the human and ecological aspects of fire risk, which can inform the development of more effective and sustainable fire management strategies. However, these studies often have the drawback of being site-specific. On the other hand, wildfire simulation models use physical models to estimate parameters and are required for fine-scale burn probabilities mapping [6,7]. By integrating spatial information from both empirical and simulation modelling, it becomes possible to take into account the interplay between human and biophysical factors that contribute to the risk of wildfires. This helps to produce a more comprehensive understanding of the drivers of wildfire risk and enables better decision making for mitigation and preparedness efforts. By combining the strengths of both empirical and simulation studies, a more holistic and accurate assessment of wildfire risk can be achieved [3].

The mapping of wildfire likelihood and potential fire hazards is used to identify forest areas with higher fire risk, where surveillance or monitoring systems can be installed. According to the most recent review studies, terrestrial systems have high accuracy and quick response times, but they have limited coverage. In order to address the limitation of single-point sensors, the deployment of extensive networks of ground sensors and the use of aerial and satellite-based systems have been proposed as alternative solutions. These systems provide improved coverage and have demonstrated high accuracy and quick response times. While fires can be detected at an early stage through terrestrial and aerial systems, satellite-based imaging sensors such as MODIS can detect small fires within 2 to 4 h after observation [2]. Given the importance of early fire detection and better coverage, this study considers the use of aerial surveillance systems.

Traditional forest fire surveillance systems capture real-time scenes, providing full-time monitoring, but blind spots always exist if the used cameras have a limited field of view (FOV). This problem can be solved by employing either a sensor network or omnidirectional cameras with a wider FOV. Early implemented methods to achieve an omnidirectional detection include the use of PTZ (pan/tilt/zoom) cameras or multiple camera clusters [8]. However, their shortages, such as the slow-moving speed of the PTZ mechanism and the additional costs of installation and maintenance, restrict their further applications. Later on, cameras with omnidirectional vision sensors such as parabolic mirror or fisheye lens have become more and more popular due to their real-time applications and low cost [9–11]. In recent years, autonomous unmanned aerial vehicles (UAVs) for early fire detection and fighting have been introduced. The goal of aerial systems is to offer a more comprehensive and precise understanding of fires from an aerial perspective. To achieve this, UAVs are utilized which integrate a variety of remote sensing technologies and computer vision techniques based on either machine learning or deep learning [12]. They mostly use ultra-high-resolution optical or infrared cameras, and they integrate various sensors for navigation and communication. More recently, 360-degree remote sensing systems have been proposed in order to capture images with unlimited field of view and early fire detection.

Traditional wildfire detection methods that utilize optical sensors have long relied on the use of various features related to the physical properties of fire, including its colour, motion, spectral, spatial, temporal, and textural characteristics. These features are then combined and analysed to accurately detect and locate wildfires [2,8]. Unlike traditional methods that utilize manually created features, deep learning techniques have the capability to automatically identify and extract intricate feature representations through the process

of learning. This results in a more sophisticated and nuanced understanding of the data, leading to improved performance in various detection and classification tasks [2,13–16]. Barmpoutis et al. [17] went a step further in utilizing deep learning methods by incorporating the concept of higher-level multidimensional texture analysis, which is based on higher-order linear dynamical systems. This combination aimed to enhance the capabilities of the deep learning approach in analysing complex patterns and structures in the data. By leveraging the strengths of both deep learning and multidimensional texture analysis, the authors aimed to achieve improved results in early fire detection. More recently, vision transformers [18–22] inspired by the deep learning model that was developed for natural language processing [23] have been employed for various applications as well as fire detection and classification of fire. More specifically, attention layers have been utilized in different ways by vision transformers. For example, Barmpoutis et al. [20] investigated the use of a spatial and multiscale feature enhancement module; Xu et al. [21] designed a fused axial attention module capturing local and global spatial interactions; and Tu et al. [22] introduced a multi-axis attention model which utilizes global–local spatial interactions. Focusing on fire detection, Ghali et al. [24] used two vision-based transformers, namely, TransUNet and MedT, extracting both global and local features in order to reduce fire-pixel misclassifications. In another study, a multi-head attention mechanism and an augmentation strategy were applied for remote sensing classification including environmental monitoring and forest detection [25]. An extended approach [26], using deep ensemble learning method and combining two vision transformers with a deep convolutional model, was employed to initially classify wildfires using aerial images and then to segment wildfire regions. Other researchers focused on the difference between clouds and smoke. More specifically, Li et al. [27] used an attention model, recursive bidirectional feature pyramid network (RBIFPN), as a backbone network of YOLOV5 frame, improving the detection accuracy of the wildfire smoke. These methods have significantly contributed to the increment of fire detection accuracy.

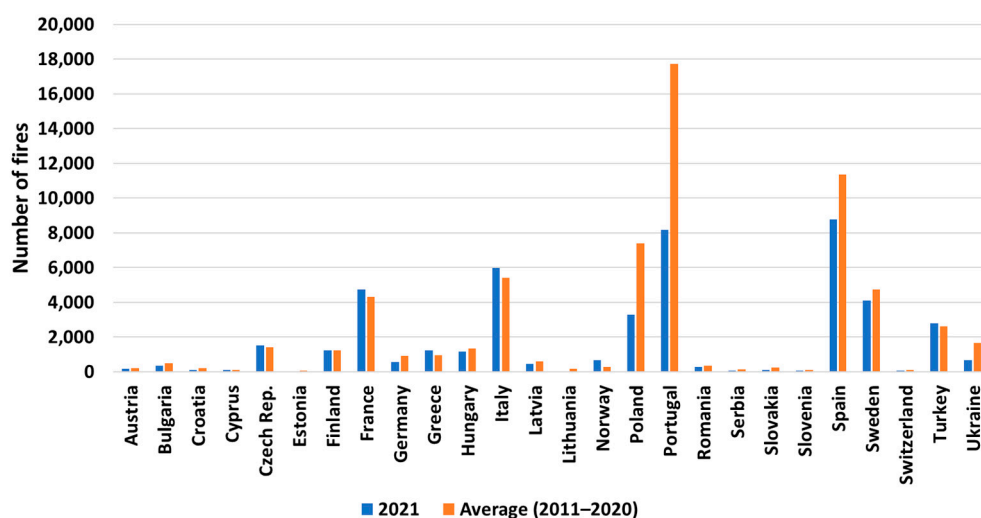
In this paper, given the urgent priority around protecting forest ecosystems, 360-degree sensors mounted to UAVs and visual transformers are used for early fire detection. More specifically, this paper makes the following contributions:

- Fire risk assessment for a Mediterranean suburban forest named Seich Sou is performed.
- A model (FIRE-mDT) that combines ResNet-50 and a multiscale deformable transformer model is adapted and introduced for early and accurate fire detection as well as for fire location and propagation estimation.
- A dataset that consists of 60 images of a real fire event which occurred on 13 July 2021 in the suburban Seich Sou Forest was created. This case was used for the evaluation of the efficiency of the real-scenario fire detection.
- Further validation was performed using the “Fire detection 360-degree dataset” [17].

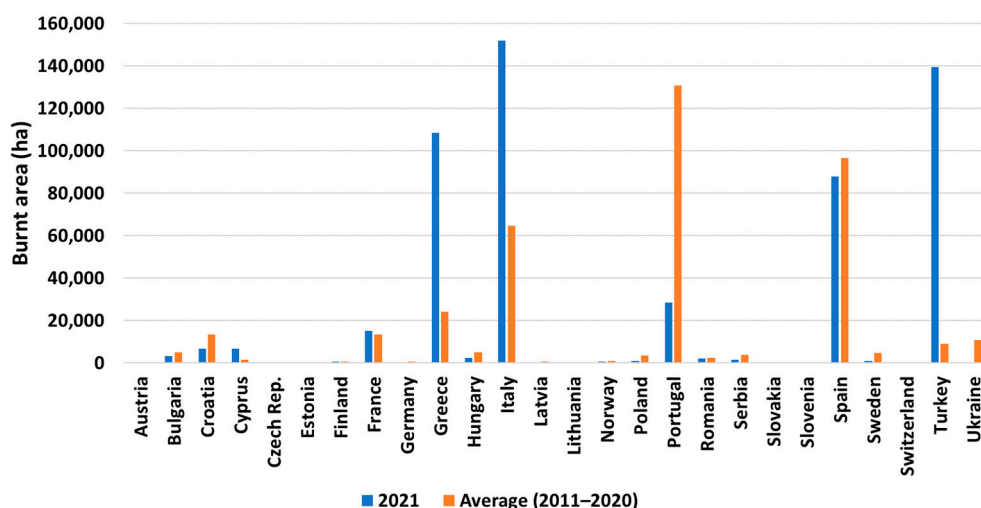
## 2. Mediterranean Suburban Fires

The current trend of climate change in the Mediterranean region that causes more intense summer season droughts and more frequent occurrence of extreme weather events has resulted in an increased number of annual forest fires in the Mediterranean region during the last decades. More specifically, fire frequency has doubled over the last century, in terms of both the number (Figure 1) of fires and burned area (Figure 2) [28]. The comparison of the number of fires for 2021 with the average number of fires for the years 2011–2020 reveals that despite the fact that the number of fires that took place in the Iberian Peninsula countries in 2021 was less than the average for the years 2011–2020, there is still a noticeable trend of increased fire occurrences and extent of burn areas in the Mediterranean countries. This highlights the need for continued vigilance and efforts to mitigate the risk of wildfires in these regions as well as the need for the development of effective early fire detection systems [29]. It is worth mentioning that 2021 was a particularly devastating year for forest fires, with many experts attributing the severity of the fires to a combination of

climate change, extreme weather events, and the impacts of the COVID-19 pandemic on land management practices.



**Figure 1.** Number of fires in EU for 2021 and the average number of fires during the years 2011–2020.

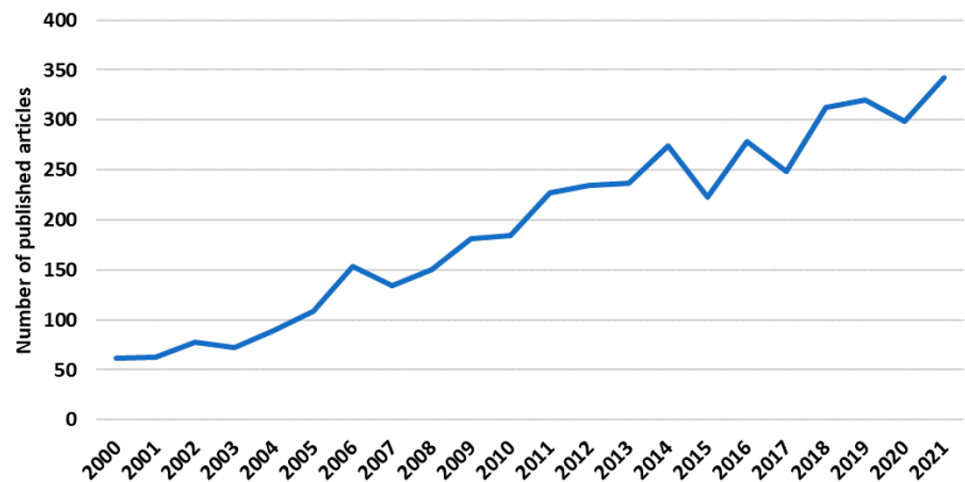


**Figure 2.** Land burned (ha) by wildfires in EU for 2021 and the average land burned (ha) by wildfires during the years 2011–2020.

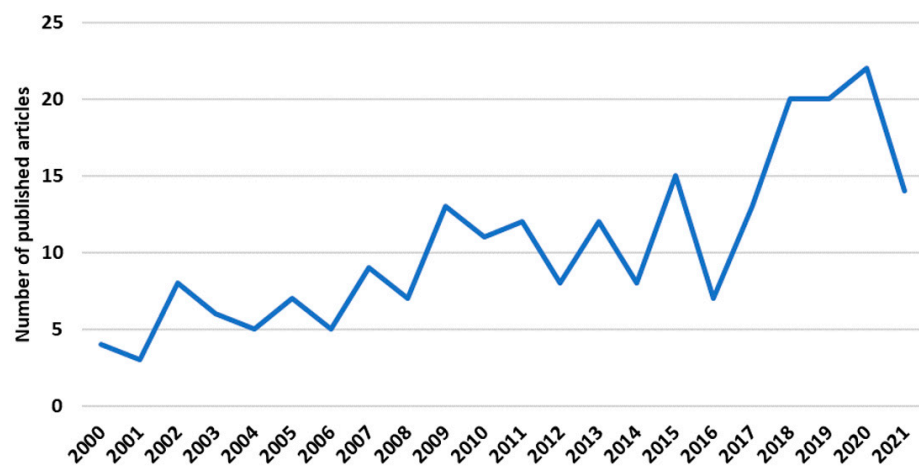
An analysis of forest fire risk in the Mediterranean basin shows that there will be a notable increase in the number of weeks with fire risk in all Mediterranean land areas between 2030 and 2060. This increase is estimated to be between 2 to 6 weeks, with a significant proportion of the increase being classified as extreme fire risk. This means that there will be a heightened likelihood of more intense and destructive fires in the region, which could have far-reaching impacts on both the environment and local communities [30]. In addition, according to a 2022 study [31], instances of extreme wildfires are forecasted to increase by 14 percent by the year 2030 and by 30 percent by 2050.

In addition, to explore the evolution of research on forest and suburban fires in the Mediterranean region, we conducted a bibliometric study. Figures 3 and 4 show the trend in the number of articles published between 2000 and 2021. The findings reveal an increase in the number of publications for both fields over the past 20 years. However, the number of published articles in the field of Mediterranean fires is higher than these in the field of suburban fires. Narrowing the results of Mediterranean fires to only the suburban fires research area, the search yielded just 16 published articles from 2000. Thus, although

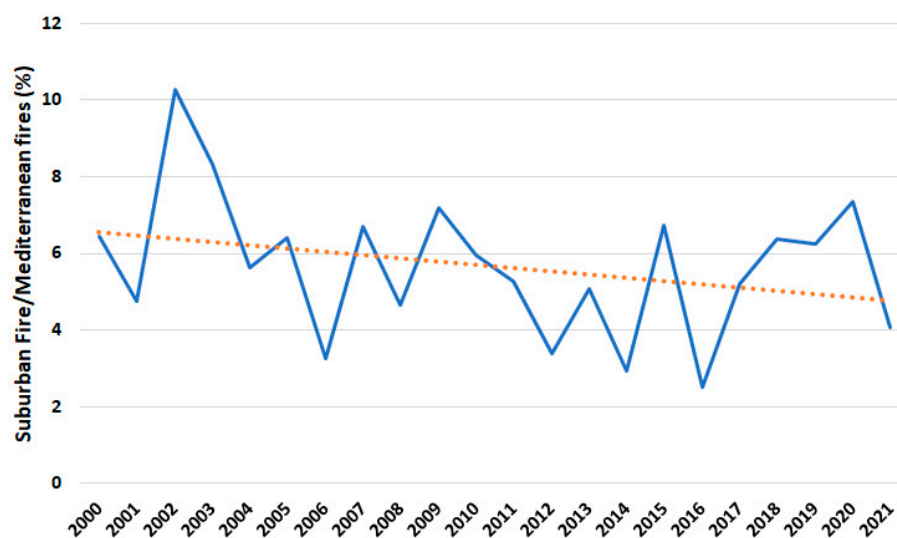
suburban forest fires pose a serious threat to both urban and rural communities, as the combination of dense vegetation and close proximity to homes and businesses can lead to rapid spread and significant damage, the bibliometric results indicate that the research in the Mediterranean fire and suburban fire areas is still evolving. In addition, Figure 5, which represents the ratio of the number of articles published related to suburban fires to the number of articles published related to Mediterranean fires, indicates that the field of Mediterranean suburban fires research has been insufficiently investigated.



**Figure 3.** The annual number of articles published related to Mediterranean fires. Data retrieved from Web of Science [32] for dates between 2000 and 2021.



**Figure 4.** The annual number of articles published related to suburban fires. Data retrieved from Web of Science [32] for dates between 2000 and 2021.



**Figure 5.** The ratio (%) of the number of articles published related to suburban fires to the number of articles published related to Mediterranean fires. Red line represents the trendline. Data retrieved from Web of Science [32] for dates between 2000 and 2021.

### 3. Materials and Methods

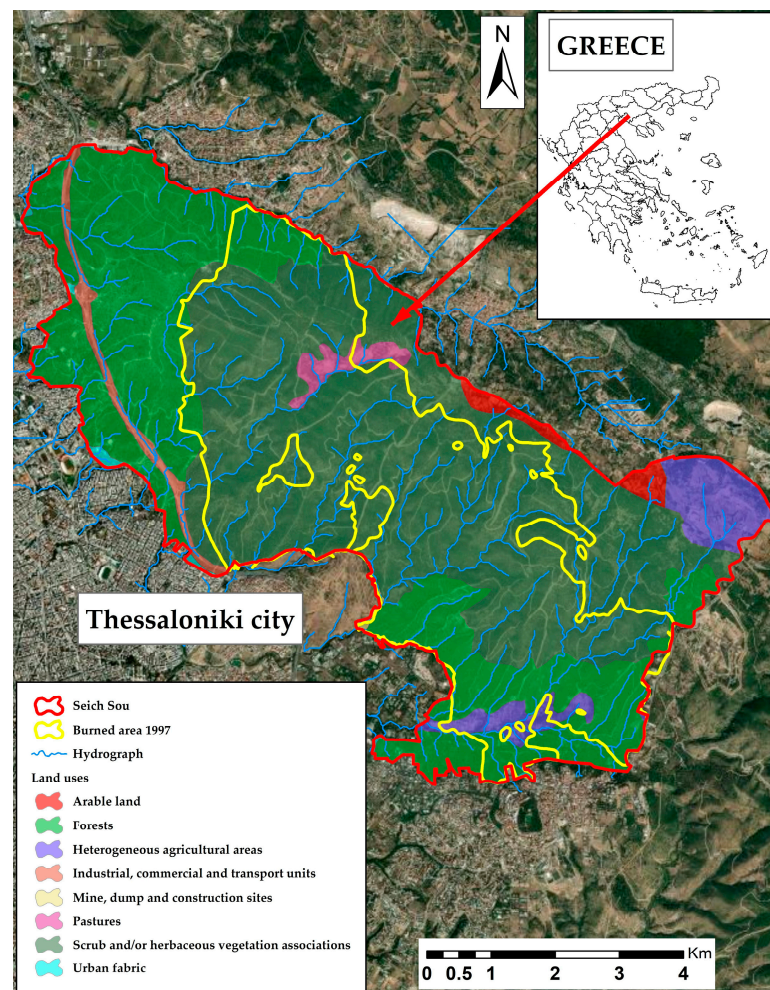
The focus of this study was the suburban forest located in Thessaloniki city, known as Seich Sou. This forest is considered one of the most significant suburban forests in Greece, encompassing a diverse range of plant and animal life. However, in 1997, the forest experienced a devastating wildfire that destroyed more than half of its area. The impact of the 1997 wildfire has been far-reaching, affecting not only the forest ecosystem but also the local community [33]. The data used in the present study were captured from various sites in the Thessaloniki metropolitan area. Initially, a fire risk assessment was performed and then a UAV equipped with 360-degree cameras was used for the forest surveillance.

#### 3.1. Suburban Forest Fire Risk Assessment

The assessment of wildfire risk requires the use of Geographic Information System (GIS) technology or aerial monitoring [34] to create, manage, and provide datasets related to fuels, land cover, and topography [35]. To model fire behaviour, raster datasets such as canopy cover, elevation, fuel, slope, and aspect are also necessary. Additionally, stand-scale features such as crown bulk density, stand height, and canopy base height are needed to characterize the canopy. Meteorological and fuel moisture data, as well as an ignition prediction raster, are also crucial components. Some of these data, such as topography and weather conditions, can be easily obtained from public databases. However, others, such as canopy structure and fuels, require more advanced methods such as remote sensing (LiDAR), spatial data analysis, and statistical tools, as well as ground observations from forest recordings, to be generated with high accuracy [36].

##### 3.1.1. Vegetation

Seich Sou forest may be classified into the following three primary forest classes based on the examination of land cover: (1) *Pinus brutia* forest from previous replanting, (2) evergreen *Ostryo-Carpinion* (pseudomaquis) vegetation, and (3) the burned portion of the forest with vegetation that was developed following the 1997 fire (Figure 6). The land cover in the area is primarily forest land (89.7%), followed by crop land (6.6%) and pastures (1.3%). The remaining 2.4% consists of the ring road and a few other special areas. The average age of “Seich Sou” is about 70 years, and the dominant tree species is *Pinus brutia* L.



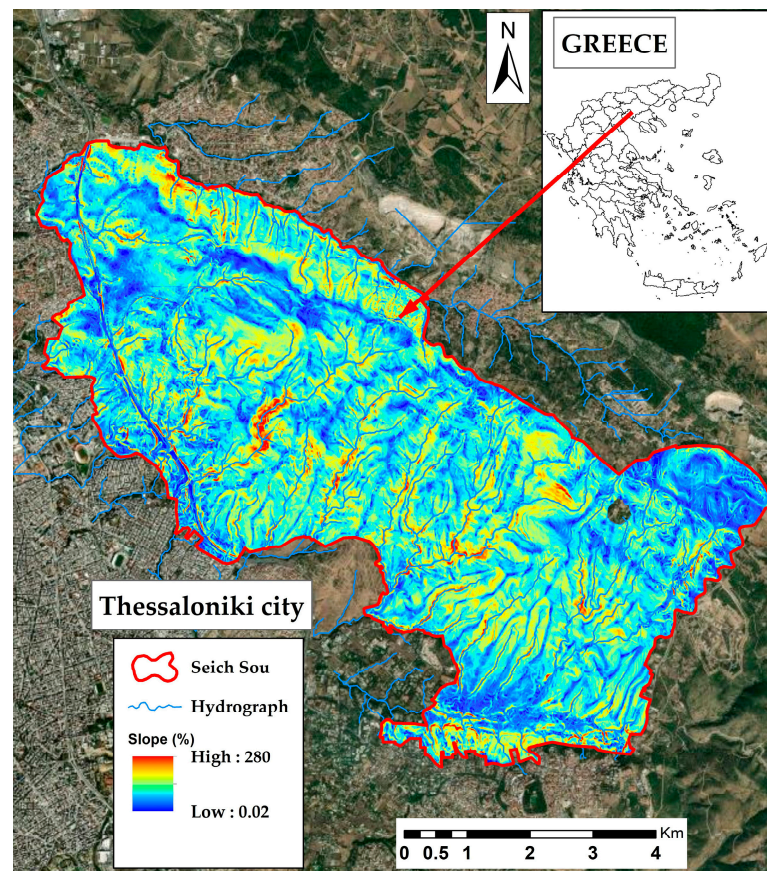
**Figure 6.** Land uses in the study area according to the Corine Land Cover 2018.

### 3.1.2. Distance from Buildings and Urban Fabric

The forest has very long boundaries with the city of Thessaloniki, a fact which significantly increases the threat of cross-boundary fire transmission.

### 3.1.3. Topography

Datasets of altitude, slope, and aspect are used to describe topography. The adiabatic alteration of temperature and humidity is regulated by the elevation. Along with aspect, the slope dataset is required for calculating the direct impacts on fire spread, adjusting fuel moisture, and converting spread velocities and directions. With a mean elevation of 306.3 m above sea level, a maximum elevation of 569.5 m, and a minimum elevation of 56.8 m, the forest's relief can be described as hilly–semi-mountainous and relatively steep, with a mean slope of 26.2% and a dominant aspect of SE–S–SW (Figure 7).



**Figure 7.** The topography and the dominant slopes of the study area.

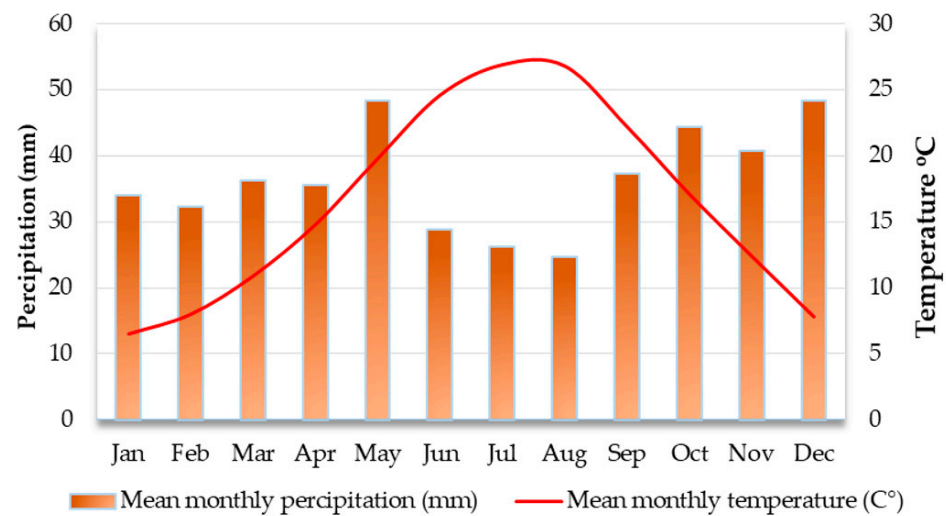
#### 3.1.4. Surface Fuel

Grasslands and forest stands burn differently across the elevation gradient. Vegetation density, height, mixture of grass, trees, and shrubs, and dead vegetation fuel highly influence the probability of ignition and the spread of fire. The vegetation of the forest is mainly of low moisture. Furthermore, the *Tomicus piniperda* bark beetle has been responsible for a severe insect infestation since May 2019 [37], which was discovered as a result of the necrosis of several pine trees (*Pinus bruttia* L.). More than 300 hectares of forest were lost due to the infestation, which is still active today. To decrease and stop the spread of the infestation, extensive selective logging has been carried out lately, as well as the removal of affected trees from the forest. Many dead trees are still present in the forest, increasing the dead fuel vegetation.

#### 3.1.5. Weather Data

The climate in the region can be characterized as a typical Mediterranean climate, which is characterized by hot and dry summers, and mild winters. This description is based on the available meteorological data, which provides a comprehensive understanding of the area's climate patterns (Figure 8). The average annual rainfall in the area is 444.5 mm, with the highest levels occurring in winter (December) and a secondary peak in spring (May). The average annual temperature is 15.9 °C, with average maximum temperature of 20.4 °C and average minimum temperature of 10.1 °C. The dominant wind direction during the summer (June, July, August) is N–NW, while the wind speed is >3.4 m/s for 33 days during the summer. The above weather conditions during the summer period increase the fire risk.





**Figure 8.** The climate in Thessaloniki, as depicted by the Bangouls and Gausson rain–temperature diagram, is characterized by a dry thermal period that starts in June and lasts until early October. This pattern represents the typical weather conditions in the region and helps to provide an understanding of the area’s climate dynamics.

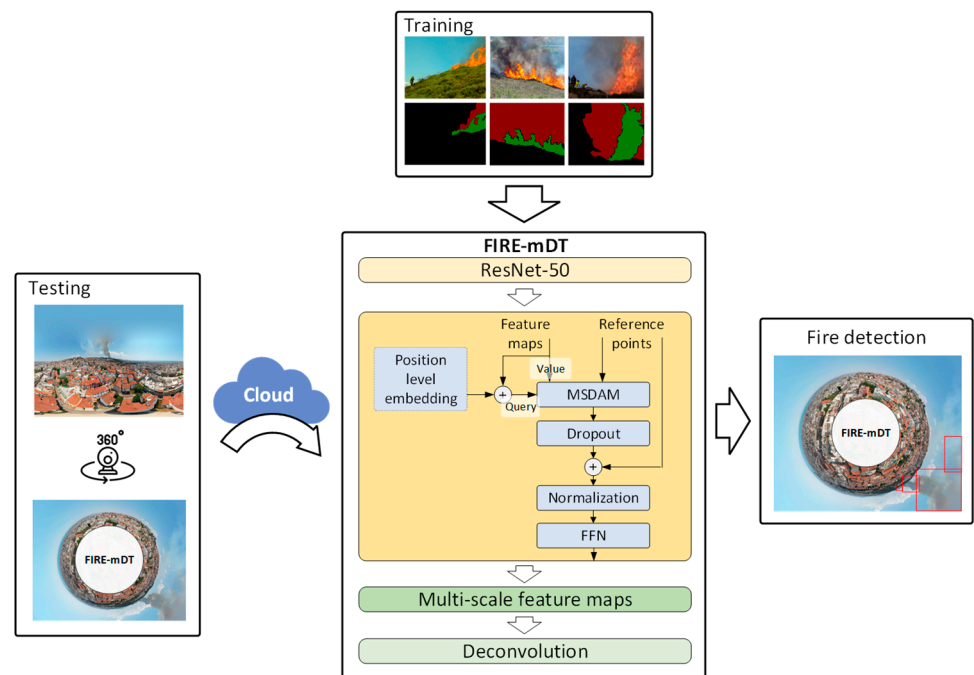
To conclude, the above fire risk parameters that were analysed, as well as other studies [38], suggest that Seich Sou, one of the most important suburban forests in Greece, is prone to fires. The rapid spread of fires in suburban forest areas can result in significant damage to both property and wildlife. Thus, the implementation of a machine-learning-based early fire detection system in suburban forests can provide the necessary surveillance for prompt fire detection and contribute to the efforts to reduce the harm caused by forest fires. Furthermore, early fire detection is crucial for reducing the costs associated with wildfire suppression (Table 1), as firefighting resources can be mobilized quickly, preventing the fire from spreading and becoming more difficult and expensive to manage.

**Table 1.** Operating costs, by type, for wildfire suppression in the Mediterranean, 2021 [39].

Type	Cost (EUR/Hour)
Airplane Canadair	8000
Helicopter S64 F	3600
Helicopter AB 412	2500
Helicopter NH 500	700
Team with heavy equipped vehicle	180
Heavy team with unequipped vehicle	150
Light team with unequipped vehicle	130
Helicopter transported team with firefighting module	105
Team with light equipped vehicle	100
Helicopter transported light team	90

### 3.2. Forest Surveillance Using Omnidirectional Cameras and Deep Vision Transformers

The proposed model (FIRE-mDT) aiming to detect early fire through 360-degree data combines the ResNet-50 as the backbone and the deformable transformer encoder-based feature extraction module [40] (Figure 9). The proposed model was applied to stereographic projections as these perform well at minimizing distortions.



**Figure 9.** The proposed methodology for early fire detection.

The detection transformer (DETR) [41] is a cutting-edge end-to-end detector; however, it requires significant amount of memory both for training and use in real-time scenarios. More specifically, the DETR has a large memory footprint due to its composition of a convolutional backbone, six encoders and decoders, and a prediction head, as well as due to the need to store the self-attention weights within each multihead self-attention layer. Therefore, for more efficient early fire detection, the deformable DETR is preferred over the original detection transformer [42]. The deformable DETR has a more efficient approach in determining attention weights as it only calculates them at specific sampling locations instead of every pixel of the feature map. This permits the training and the testing of the model with high-resolution data using standard GPUs, making it more accessible and convenient to use. Furthermore, its performance has been demonstrated to be superior to the Faster R-CNN through extensive experiments [43], making the deformable DETR a promising solution for early fire detection.

The FIRE-mDT model is designed to retain high-level semantic information and maintain feature resolution, which is achieved through the combination of the ResNet-50 backbone and the deformable transformer module. By setting the stride and dilation of the final stage of the backbone to 1 and 2, respectively, the FIRE-mDT is able to extract higher-level semantic information. The final three feature maps are then fed into the deformable transformer encoder, with the first two being upsampled by a factor of two and the third being encoded through a convolutional layer. The multiscale feature maps are then concatenated, group normalized, and are finally inputted into the deformable transformer encoder for multiscale feature extraction. The result of this process is the multiscale deformable attention feature map, which incorporates the adaptive spatial features that have been added through the self-attention mechanism in the encoder, providing a robust and accurate solution for early fire detection. The model takes into account both local and global dependencies, which are enhanced through the self-attention mechanism.

The deformable transformer encoder, a key component of the FIRE-mDT model, works by enhancing the input feature maps with positional encodings and level information resulting the query vector  $z_q$ . The  $z_q$  is then used as input, along with the feature maps and reference points, into the multiscale deformable attention module (MSDAM). Then, the MSDAM extracts a multiscale, deformable attention feature map. To generate this map, the MSDAM initially computes value, weight, and location tensors and then employs them in

the multiscale deformable attention function. The  $q$ -th element of the deformable attention feature  $z' \in \mathbb{R}^{N_q \times c_v}$  ( $N_q = \sum_{l=1}^3 H_l W_l$ ) at a single head is expressed as follows:

$$z'_q = \sum_p \sum_{l=1}^3 W_{plhq} v_{p_{ql} + \Delta p_{qhlp}} \quad (1)$$

where  $q, h$ , and  $p$  represents the components of the output of the deformable attention feature  $z_o$ , the attention head, and the sampling offsets, respectively.  $W_{plhq}$  is an entity of  $W \in \mathbb{R}^{N_q \times N_h \times 3 \times N_p}$ . Furthermore,  $p_{ql}$  and  $\Delta p_{qhlp}$  denote the position of a reference point and one of the  $N_p$  corresponding sampling offsets of  $p \in \mathbb{R}^{N_q \times 3 \times 2}$  and  $\Delta p \in \mathbb{R}^{N_q \times N_h \times 3 \times N_p \times 2}$ , respectively. The number of sampling offsets and attention head are set as  $N_p = 4$  and  $N_h = 8$ . Afterwards, the input feature maps are combined with the deformable attention feature map, and then processed through a feedforward network (FFN).

Additionally, the FIRE-mDT model includes a final step to refine the results of the deformable encoder. This step involves a layer of deconvolution and normalization followed by a process of feature fusion that merges multiple features and then passes them through a convolutional layer. The final layer of the model is a regressor that outputs the fire detection results. Finally, to optimize the model performance, we used a focal loss [44] that defines higher weights on difficult examples, aiming to improve the precision of the model predictions. The optimization process was carried out using the Adam optimizer, along with a mean teacher method [45], which aims to improve the robustness of the model performance. The learning rate was set equal to  $0.5 \times 10^{-4}$  and decreased by a factor of 0.5 after 50 epochs. The model was trained using a single NVIDIA GeForce RTX 3090 GPU for 80 epochs with batch size of 16.

### 3.3. Dataset Description and Evaluation Metrics

For the evaluation of the proposed framework, we used two different datasets. The first dataset includes sixty 360-degree stereographic projection images of the fire event which occurred on 13 July 2021 in the suburban Seich Sou Forest in Thessaloniki. The dataset comprises fifty-eight images captured from the early stages to the late stages of the fire on 13 July 2021 and two images captured after the fire incident. In this fire incident, 90 acres of forest land were burned and 66 firefighters, 22 firefighting vehicles, three Canadair firefighting planes, and two helicopters were mobilized. This dataset was used for the evaluation of the proposed framework in a real fire scenario.

In addition, we used a publicly available dataset named “Fire detection 360-degree dataset” in order to compare the effectiveness of our proposed approach with other state-of-the-art methods. This dataset contains both synthetic and real fire data.

For the training of the proposed FIRE-mDT model, we used the Corsican Fire Database (CFDB) [46,47]. Moreover, in addition to the flame annotations in the dataset, smoke annotations were also performed in this study. In addition, to further increase the variability of the training dataset and make the model more robust, an augmentation method was applied. This involved making random modifications to the images in the training set, such as rotation, flipping, and scaling, to increase the diversity of the training data and help the model better generalize to different conditions and scenarios.

In order to gauge the effectiveness of the proposed fire detection model, two evaluation metrics were employed: *F-score* [48] and mean intersection over union (*mIoU*) [49]. The intersection over union, also known as Jaccard index, is a well-established metric that provides a quantitative measurement of the overlap between the detected fire region and the ground truth. It is calculated by dividing the area of overlap between the two regions by the area of their union. This metric is useful in determining the accuracy of fire detection

and provides a comprehensive understanding of how well the model is able to detect fires in an image. More specifically, the *mIoU* is defined as follows:

$$IoU = \frac{groundTruth \cap prediction}{groundTruth \cup prediction} \quad (2)$$

The *F-score* is a widely adopted evaluation metric that assesses the accuracy of a fire detection model. It is calculated as the harmonic mean of precision and recall, which are two important parameters that measure the performance of a fire detection system. Precision refers to the proportion of correctly identified fire regions among all the detected regions, while recall represents the proportion of correctly identified fire regions among all the actual fire regions. For the calculation of the precision and recall, we estimated the number of correctly detected fire images, the number of false negative images, and the number of images that includes false positives. The *F-score* is important because it considers both precision and recall, providing a more comprehensive evaluation of the fire detection model's performance. By combining these two metrics, the *F-score* provides an overall assessment of the model's accuracy and helps to identify areas where improvement may be necessary. More specifically, the *F-score* is defined as follows:

$$Fscore = 2 \frac{precision \cdot recall}{precision + recall} \quad (3)$$

#### 4. Results and Discussion

The evaluation of the proposed fire detection methodology is presented in this section. The experimental evaluation has two main objectives: first, to showcase the performance results of the fire detection model during the Thessaloniki fire event on July 13 2021; second, to showcase the advantage of the proposed framework using 360-degree data over state-of-the-art approaches.

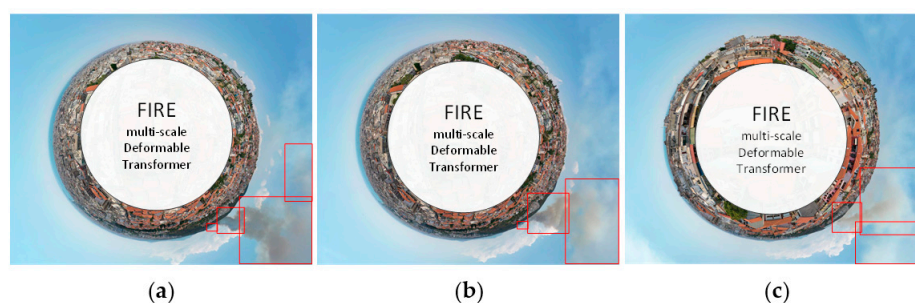
##### 4.1. Case Study of the "Seich Sou" Suburban Forest of Thessaloniki

This study took place in the Seich Sou suburban forest of Thessaloniki city, and a fire detection study of the fire event which occurred on 13 July 2021 (Figure 10) was performed. For the evaluation of the proposed framework, we created a 360-degree dataset, consisting of sixty 360-degree equirectangular images of the suburban forest of the Thessaloniki fire event. In a similar manner to the previous study [17], we utilized a 360-degree camera equipped with GPS mounted on an unmanned aerial vehicle (UAV) for capturing 360-degree images. The camera used in this study had a CMOS sensor type with a 1/2.3" sensor size.



Figure 10. Fire event captured in Thessaloniki on 13 July 2021.

The proposed model was applied to the 360-degree images of the fire event in Seich Sou, resulting in an  $F$ -score of 91.6%. More specifically, to estimate the  $F$ -score and to evaluate the effectiveness of the proposed fire detection algorithm, we estimated the number of correctly detected images, identified by at least one accurately detected fire region (true positive) and the number of images that correctly classified as negative to fire (true negative), as well as the number of missed fire images (false negative) and the images that erroneously contained at least one positive fire region (false positive). The proposed model achieved forty-nine true positive images (Figure 11), correctly identifying regions of fire, as well as two true negative images, accurately identifying regions without fire. However, there were also five false negative images, where the model failed to detect the presence of fire (Figure 12). In addition, four images were found to have at least one false positive fire region detected in a nonfire region. In general, the framework that was proposed appears to be strong enough to effectively handle both false negatives and false alarms, indicating a high potential for practical use. These results highlight both the strengths and weaknesses of the proposed model, with a relatively high true positive rate and four false positive images indicating a high potential for practical use.



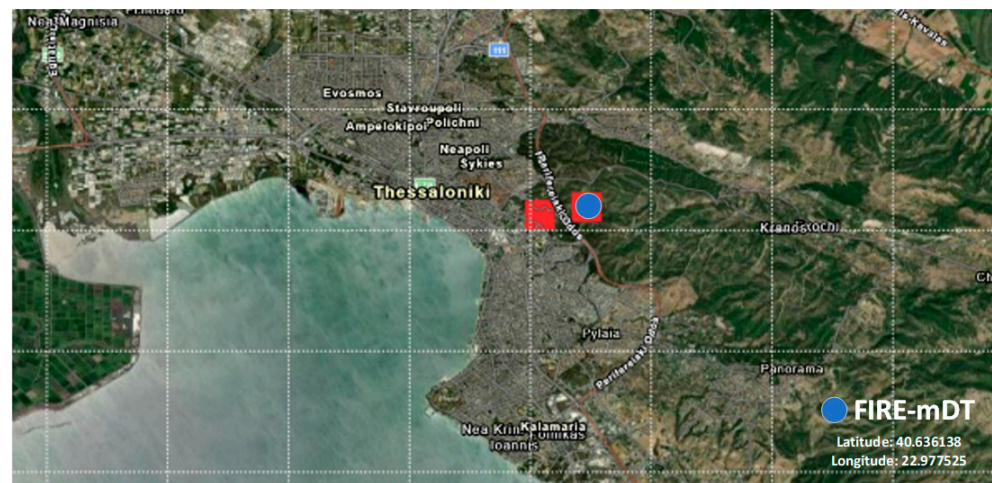
**Figure 11.** Fire detection (red boxes) of the FIRE-multiscale deformable transformer applied to 360-degree stereographic projections at three different time points (a–c) of the fire event.



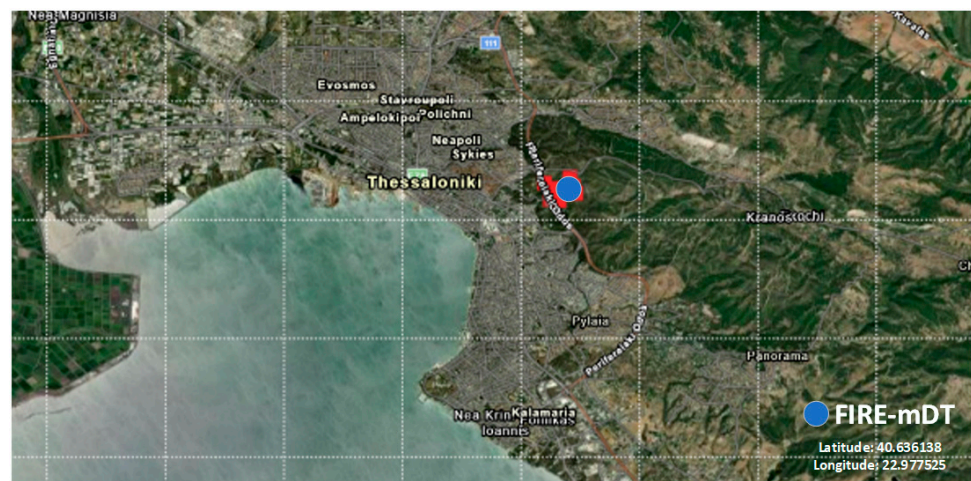
**Figure 12.** Fire misdetection of the FIRE-multiscale deformable transformer applied to 360-degree stereographic projections.

In addition, we investigated the effectiveness of multiscale deformable attention feature maps in fire detection. To this end, we compared the utilization of a single deformable attention feature map with that of multiple deformable attention feature maps. Our analysis revealed that the use of multiscale feature maps resulted in a slight improvement in the  $F1$ -score, with an increase of 1.4. The outcomes further demonstrated that higher-level features attained better precision, while lower-level features obtained a better recall score.

Furthermore, taking into account the altitude of the 360-degree camera (in the range of 18 m to 28 m) as well as the GPS coordinates of the UAV (latitude and longitude), the location of the fire was easily estimated, allowing for fire management and planning (Figures 13 and 14). In addition, capturing consecutive images also enables the estimation of fire propagation. In this fire incident, the fire services were able to quickly contain the fire and prevent it from spreading significantly due to the combination of low wind conditions and their prompt response. As a result, the fire was successfully controlled and did not cause any major propagation.



**Figure 13.** Fire data through MODIS (red boxes) and fire-mDT detected location (blue circle) for the fire incident on 13 July 2021.



**Figure 14.** Fire data through VIIRS (red boxes) and fire-mDT detected location (blue circle) for the fire incident on 13 July 2021.

Furthermore, for the analysis of this fire event, Fire Information for Resource Management System (FIRMS) was used for the analysis of MODIS and visible infrared imaging radiometer suite (VIIRS) data. It is worth mentioning that the MODIS captures data in 36 spectral bands while the VIIRS provides 22 different spectral bands. Based on the analysis of the suburban fire on 13 July 2021, both MODIS (Figure 13) and VIIRS (Figure 14) did not precisely locate the fire as they identified that there were flames on both sides of the ring road of Thessaloniki. In actuality, the fire started only on the upper side of the road. This mislocalization could significantly affect the operational fire management and rescue planning. In terms of latency time, the MODIS fire products are typically generated and delivered to fire management partners within 2–4 h of the MODIS data collection under nearly optimal conditions. However, the terrestrial and aerial-based systems can detect fires at a very early stage with a much shorter latency time, providing fire management teams with critical information in a timely manner to respond quickly to fire incidents. These systems offer a significant advantage over MODIS in terms of speed and effectiveness in detecting early signs of a fire, allowing fire management teams to take proactive measures to prevent the spread of fires.

#### 4.2. Comparison Evaluation

In this study, the proposed fire detection algorithm was thoroughly evaluated using the “Fire detection 360-degree dataset” [17]. This dataset comprises 150 360-degree images of both forest and urban areas, including both artificially generated and real instances of fire events. To compare the performance of the proposed framework with other state-of-the-art methods, the evaluation results of the proposed framework were analysed and compared with seven other methods. The results of the comparison are presented in Table 2, which provides a comprehensive evaluation of the proposed algorithm’s performance. The proposed method is compared against SSD [50], FireNet [51], YOLO v3 [52], Faster R-CNN [43], Faster R-CNN/gVLAD encoding [13], U-Net [53], and DeepLab [17] architectures, and it outperforms these, mainly due to the integrated deformable transformer encoder. More specifically, the proposed system towards fire detection achieves *F-score* rates of 95%, improving by up to 0.6% the fire detection results against the second-best approach and the combination of two DeepLab v3+ and an adaptive post-validation method. It is worth mentioning that, in contrast to the DeepLab approach [17], the proposed model is a single step end-to-end approach. Similarly, the proposed model achieved an *mIoU* rate of 78.4% versus the second-best approach, which achieved an *mIoU* equal to 77.1%.

**Table 2.** Comparison results.

Method	Fire Detection	
	mIoU	F-Score
SSD [50]	59.8	67.6
FireNet [51]	61.4	71.1
YOLO v3 [52]	69.5	78.8
Faster R-CNN [43]	65.0	71.5
Faster R-CNN/gVLAD encoding [13]	73.8	87.4
U-Net [53]	67.4	71.9
2×DeepLab v3+ and an adaptive post-validation method [17]	77.1	94.6
Proposed/FIRE-multiscale deformable transformer	78.4	95

#### 4.3. Discussion

During the last decades, urban and wildland regions that cross or intermingle have significantly increased around the world, and these pose a high danger of forest fires. An environment at risk of fire where property and human lives are directly threatened is produced by the interaction of many anthropogenic agents of ignition and combustible forest vegetation with human infrastructure. Thus, both forest fire risk assessment and monitoring approaches need to be combined for an effective prevention and early detection of forest fires. The identification of forests at risk and the ideal locations for the installation of fire detection sensors could be aided by fire risk assessment, while forest detection contributes to the protection of forests and significantly reduces the amount of forest land that is destroyed by fires.

Compared to traditional narrow field-of-view sensors, omnidirectional sensors provide a broader field of view from single or several viewpoints so that they can supply richer visual information of recording areas without suffering from blind spots. Wide field of view or omnidirectional images can be acquired by rotating cameras or employing multiple camera clusters. The former are limited to obtaining omnidirectional images of static scenes; therefore, they cannot be used for real-time applications.

An omnidirectional camera can be used in real-time applications, so it has the advantage of solving any security problems in real time. In addition, cost savings on installation and maintenance can be achieved by replacing multiple narrow field-of-view cameras with a single omnidirectional camera. Although multiple cameras can provide relatively high-resolution omnidirectional images, thanks to the development of higher resolutions and storage technology for video data, large coverage and rich details become more affordable using 360-degree sensors.

Recently, 360-degree optical cameras have been suggested as a flexible and cost-effective remote sensing option for early fire detection. In this study, a vision transformer is introduced as a solution for early fire detection. In addition, we extended the evaluation of their application, and for the first-time, omnidirectional sensors were applied to a real forest fire scenario. The proposed single-stage model seems to be robust, indicating that the proposed framework has the potential to significantly enhance early forest fire detection. Based on the findings of the current study and the case study of the fire event of Seich Sou on 13 July 2021, the application of the proposed framework resulted in an *F-score* equal to 91.6%. The false negatives in the model's performance can be partly attributed to the fact that the forest fire had been extinguished and was inactive during the late image acquisition process. As a result, some regions that were previously on fire appeared as nonfire regions in the images, leading to false negatives. On the other hand, some nonfire regions, such as clouds or bright areas, may have been mistakenly identified as fire regions, leading to false positives. Additionally, the thinness of the smoke produced by the fire also contributed to the model's difficulty in detecting the fire regions accurately, resulting in both false negatives and false positives. It is important to mention that the fire detection model proposed in this study shows improved performance compared to other advanced methods applied to a dataset that is publicly available. This highlights the effectiveness and potential of the proposed approach in the field of early fire detection.

On the other hand, deploying and maintaining a 360-degree camera mounted on a UAV for fire surveillance can involve various potential costs and challenges. The initial cost of acquiring and maintaining the necessary equipment, including the UAV itself, the 360-degree camera, sensors, software, and other accessories, can be high. The UAV's limited battery life can also pose a challenge for extended surveillance missions, requiring multiple batteries and flights for extended periods of time. In addition, weather conditions such as high winds, rain, and snow can affect the operation of the UAV and the 360-degree camera system, necessitating additional maintenance. Furthermore, operating the UAV system involves inherent safety risks, which can cause damage to the equipment or pose a danger to personnel or the public. This highlights the need for personnel operating or maintaining the UAV system to receive specialized training and certification. Additionally, the data collected from the 360-degree camera must be stored and analysed, which may require additional resources and infrastructure. Furthermore, the UAV must be regularly serviced and maintained to ensure that it operates at peak performance.

Finally, the proposed methodology for early fire detection, including the conversion of the equirectangular images to stereographic projections, achieves a processing speed of 5.3 frames per second (FPS). This result is considered to be good for real-time applications, as it enables the methodology to keep up with rapidly changing conditions and make quick decisions based on the information it detects. This is particularly important in the case of fire detection, where early detection can be critical in preventing or mitigating the spread of fires.

## 5. Conclusions

Forest risk assessment helps to identify and locate the vulnerable zones and can assist in early fire detection. In this study, fire risk assessment of the Seich Sou suburban forest of Thessaloniki was performed and the recently introduced 360-degree sensors were used for early fire detection. The proposed multiscale deformable transformer encoder-based detection method offers a new end-to-end solution to single-stage forest fire detection, with the ability to accurately identify fire events in 360-degree data. This approach extracts scale-aware features, emphasizing the most important features and reducing the attention on background information. The results of the proposed model on the "Fire detection 360-degree dataset" show that it is effective and robust in detecting fire events. This highlights the potential of this novel architecture to significantly improve the accuracy and reliability of fire detection systems. The application of the proposed methodology in a real fire event indicates that the proposed framework has the potential to make a significant



contribution to early fire detection in the suburban forest of Thessaloniki. In the future, we aim to use multispectral 360-degree sensors, expanding the fire detection capabilities.

**Author Contributions:** Conceptualization, P.B., A.K., T.S. and N.G.; methodology, P.B., A.K., T.S., J.Y., M.S. and N.G.; software, P.B. and A.K.; validation, P.B., A.K., T.S. and N.G.; formal analysis, P.B. and A.K.; investigation, P.B., A.K., T.S. and N.G.; resources, P.B., A.K., T.S. and N.G.; data curation, P.B., A.K., T.S. and N.G.; writing—original draft preparation, P.B. and A.K.; writing—review and editing, P.B., A.K., T.S. and N.G.; visualization, P.B. and A.K.; supervision, P.B.; project administration, P.B.; funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been funded by Greece and the European Union through the Operational Programme “Human Resources Development, Education and Lifelong Learning” in the context of the call “Reinforcement of Postdoctoral Researchers—2nd Cycle” (MIS 5033021) for the project i-FORESTER: Intelligent system for FOREST firE surveillance. Nikos Grammalidis has received funding from the INTERREG V-A COOPERATION PROGRAMME Greece-Bulgaria 2014–2020 project “e-OUTLAND: Protecting biodiversity at NATURA 2000 sites and other protected areas from natural hazards through a certified framework for cross-border education, training and support of civil protection volunteers based on innovation and new technologies.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Finney, M.A. The challenge of quantitative risk analysis for wildland fire. *For. Ecol. Manag.* **2005**, *211*, 97–108. [\[CrossRef\]](#)
2. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Ager, A.A.; Preisler, H.K.; Arca, B.; Spano, D.; Salis, M. Wildfire risk estimation in the Mediterranean area. *Environmetrics* **2014**, *25*, 384–396. [\[CrossRef\]](#)
4. Verde, J.C.; Zêzere, J. Assessment and validation of wildfire susceptibility and hazard in Portugal. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 485–497. [\[CrossRef\]](#)
5. Lautenberger, C. Mapping areas at elevated risk of large-scale structure loss using Monte Carlo simulation and wildland fire modeling. *Fire Saf. J.* **2017**, *91*, 768–775. [\[CrossRef\]](#)
6. Ager, A.A.; Vaillant, N.M.; Finney, M.A. A comparison of landscape fuel treatment strategies to mitigate wildland fire risk in the urban interface and preserve old forest structure. *For. Ecol. Manag.* **2010**, *259*, 1556–1570. [\[CrossRef\]](#)
7. Ager, A.A.; Vaillant, N.M.; Finney, M.A. Integrating Fire Behavior Models and Geospatial Analysis for Wildland Fire Risk Assessment and Fuel Management Planning. *J. Combust.* **2011**, *2011*, 1–19. [\[CrossRef\]](#)
8. Etin, A.E.; Dimitropoulos, K.; Gouverneur, B.; Grammalidis, N.; Günay, O.; Habiboğlu, Y.H.; Töreyn, B.U.; Verstockt, S. Video fire detection—Review. *Digital Signal Process.* **2013**, *23*, 1827–1843.
9. Delforouzi, A.; Grzegorzec, M. Robust and Fast Object Tracking for Challenging 360-degree Videos. In Proceedings of the 2017 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 11–13 December 2017; pp. 274–277. [\[CrossRef\]](#)
10. Liu, D.; An, P.; Ma, R.; Zhan, W.; Ai, L. Scalable omnidirectional video coding for real-time virtual reality applications. *IEEE Access* **2018**, *6*, 56323–56332. [\[CrossRef\]](#)
11. Wang, K.H.; Lai, S.H. Object detection in curved space for 360-degree camera. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3642–3646.
12. Bouguettaya, A.; Zazour, H.; Taberkit, A.M.; Kechida, A. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Process.* **2022**, *190*, 108309. [\[CrossRef\]](#)
13. Barmpoutis, P.; Dimitropoulos, K.; Kaza, K.; Grammalidis, N. Fire Detection from Images Using Faster R-CNN and Multidimensional Texture Analysis. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12 May 2019; pp. 8301–8305. [\[CrossRef\]](#)
14. Jiao, Z.; Zhang, Y.; Xin, J.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. A deep learning based forest fire detection approach using UAV and YOLOv3. In Proceedings of the 2019 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 23–27 July 2019; pp. 1–5.
15. Xue, Z.; Lin, H.; Wang, F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. *Forests* **2022**, *13*, 1332. [\[CrossRef\]](#)
16. Ulku, I.; Barmpoutis, P.; Stathaki, T.; Akagunduz, E. Comparison of Single Channel Indices for U-Net Based Segmentation of Vegetation in Satellite Images. In Proceedings of the Twelfth International Conference on Machine Vision, Amsterdam, The Netherlands, 16–18 November 2019; Volume 11433, p. 1143319. [\[CrossRef\]](#)
17. Barmpoutis, P.; Stathaki, T.; Dimitropoulos, K.; Grammalidis, N. Early Fire Detection Based on Aerial 360-Degree Sensors, Deep Convolution Neural Networks and Exploitation of Fire Dynamic Textures. *Remote Sens.* **2020**, *12*, 3177. [\[CrossRef\]](#)
18. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [\[CrossRef\]](#)

19. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [[CrossRef](#)]
20. Barmpoutis, P.; Yuan, J.; Waddingham, W.; Ross, C.; Hamzeh, K.; Stathaki, T.; Alexander, D.C.; Jansen, M. Multi-scale Deformable Transformer for the Classification of Gastric Glands: The IMGL Dataset. In Proceedings of the first international workshop, CaPTion 2022, held in conjunction with MICCAI, Singapore, 18–22 September 2022.
21. Xu, R.; Tu, Z.; Xiang, H.; Shao, W.; Zhou, B.; Ma, J. CoBEVT: Cooperative Bird’s Eye View Semantic Segmentation with Sparse Transformers. In Proceedings of the 6th Annual Conference on Robot Learning; 2022.
22. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxvit: Multi-axis vision transformer. In Proceedings of the Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Springer Nature: Cham, Switzerland, 2022; pp. 459–479.
23. Kenton, J.D.M.W.C.; Toutanova, L.K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2 June 2019; pp. 4171–4186.
24. Ghali, R.; Akhloufi, M.A.; Jmal, M.; Souidene Mseddi, W.; Attia, R. Wildfire segmentation using deep vision transformers. *Remote Sens.* **2021**, *13*, 3527. [[CrossRef](#)]
25. Bazi, Y.; Bashmal, L.; Rahhal, M.M.A.; Dayil, R.A.; Ajlan, N.A. Vision transformers for remote sensing image classification. *Remote Sens.* **2021**, *13*, 516. [[CrossRef](#)]
26. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors* **2022**, *22*, 1977. [[CrossRef](#)]
27. Li, A.; Zhao, Y.; Zheng, Z. Novel Recursive BiFPN Combining with Swin Transformer for Wildland Fire Smoke Detection. *Forests* **2022**, *13*, 2032. [[CrossRef](#)]
28. Pausas, J.G.; Fernández-Muñoz, S. Fire regime changes in the Western Mediterranean Basin: From fuel-limited to drought-driven fire regime. *Clim. Chang.* **2011**, *110*, 215–226. [[CrossRef](#)]
29. Wildfires in the Mediterranean. 2022. Available online: <https://www.statista.com/study/53771/wildfires-in-the-mediterranean/> (accessed on 20 January 2023).
30. Giannakopoulos, C.; Le Sager, P.; Bindi, M.; Moriondo, M.; Kostopoulou, E.; Goodess, C. Climatic changes and associated impacts in the Mediterranean resulting from a 2 °C global warming. *Glob. Planet. Chang.* **2009**, *68*, 209–224. [[CrossRef](#)]
31. Sönnichsen, N. Global Estimated Wildfire Growth 2030–2050. Available online: <https://www.statista.com/statistics/1292700/global-forecast-increase-in-wildfires/> (accessed on 20 January 2023).
32. Institute for Scientific Information (ISI) Web of Knowledge/Science. Available online: <https://apps.webofknowledge.com> (accessed on 6 March 2019).
33. Margiorou, S.; Kastridis, A.; Sapountzis, M. Pre/Post-Fire Soil Erosion and Evaluation of Check-Dams Effectiveness in Mediterranean Suburban Catchments Based on Field Measurements and Modeling. *Land* **2022**, *11*, 1705. [[CrossRef](#)]
34. Barmpoutis, P.; Stathaki, T.; Kamperidou, V. Monitoring of Trees’ Health Condition Using a UAV Equipped with Low-cost Digital Camera. *ICASSP IEEE Int. Conf. Acoust. Speech Signal Process. Proc.* **2019**, *2019*, 8291–8295. [[CrossRef](#)]
35. Nuthammachot, N.; Stratoulis, D. Multi-criteria decision analysis for forest fire risk assessment by coupling AHP and GIS: Method and case study. *Environ. Dev. Sustain.* **2021**, *23*, 17443–17458. [[CrossRef](#)]
36. Myroniuk, V.; Kutia, M.; JSarkissian, A.; Bilous, A.; Liu, S. Regional-scale forest mapping over fragmented land-scapes using global forest products and Landsat time series classification. *Remote Sens.* **2020**, *12*, 187. [[CrossRef](#)]
37. Barmpoutis, P.; Kamperidou, V.; Stathaki, T. Estimation of extent of trees and biomass infestation of the suburban forest of Thessaloniki (Seih Sou) using UAV imagery and combining R-CNNs and multichannel texture analysis. In Proceedings of the Twelfth International Conference on Machine Vision (ICMV 2019), Amsterdam, The Netherlands, 16–18 November 2020; Volume 11433, pp. 910–917. [[CrossRef](#)]
38. Siachalou, S.; Doxani, G.; Tsakiri-Strati, M. Integrating remote sensing processing and GIS to fire risk zone mapping: A case study for the Seih Sou forest of Thessaloniki. In Proceedings of the International Cartography Conference (ICC), Santiago, Chile, 15–21 November 2009.
39. Alves, B. Estimated Operating Costs for Wildfire Suppression in the Mediterranean in 2021, by Type. 2009. Available online: <https://www.statista.com/statistics/1296801/operating-costs-to-suppress-wildfires-in-the-mediterranean/> (accessed on 20 January 2023).
40. Yuan, J.; Barmpoutis, P.; Stathaki, T. Effectiveness of Vision Transformer for Fast and Accurate Single-Stage Pedestrian Detection. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27427–27440.
41. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Volume 12346, pp. 213–229. [[CrossRef](#)]
42. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
43. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28: 29th Annual Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.

44. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
45. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
46. Corsican Fire Database. Available online: <http://cfd.db.univ-corse.fr/modules.php?name=Sections&sop=viewarticle&artid=137&menu=3> (accessed on 8 September 2022).
47. Toulouse, T.; Rossi, L.; Campana, A.; Celik, T.; Akhloufi, M.A. Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Saf. J.* **2017**, *92*, 188–194. [[CrossRef](#)]
48. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *AI 2006: Advances in Artificial Intelligence, 19th Australian Joint Conference on Artificial Intelligence 2006, Hobart, Australia, 4–8 December 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1015–1021.
49. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
50. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
51. Jadon, A.; Omama, M.; Varshney, A.; Ansari, M.S.; Sharma, R. Firenet: A specialized lightweight fire & smoke detection model for real-time iot applications. *arXiv* **2019**, arXiv:1905.11922.
52. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
53. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.