Article

Towards a Deep-Learning-Based Framework of Sentinel-2 Imagery for Automated Active Fire Detection

Qi Zhang [1,2], Linlin Ge [2], Ruiheng Zhang [1,*] , Graciela Isabel Metternicht [3] , Chang Liu [2] and Zheyuan Du [2]

1   School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China; zhang.qi@geos.org.au
2   School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia; l.ge@unsw.edu.au (L.G.); chang.liu17@student.unsw.edu.au (C.L.); z.du@unsw.edu.au (Z.D.)
3   School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia; g.metternicht@unsw.edu.au
*   Correspondence: ruiheng.zhang@bit.edu.cn

Abstract: This paper proposes an automated active fire detection framework using Sentinel-2 imagery. The framework is made up of three basic parts including data collection and preprocessing, deep-learning-based active fire detection, and final product generation modules. The active fire detection module is developed on a specifically designed dual-domain channel-position attention (DCPA)+HRNetV2 model and a dataset with semi-manually annotated active fire samples is constructed over wildfires that commenced on the east coast of Australia and the west coast of the United States in 2019–2020 for the training process. This dataset can be used as a benchmark for other deep-learning-based algorithms to improve active fire detection accuracy. The performance of active fire detection is evaluated regarding the detection accuracy of deep-learning-based models and the processing efficiency of the whole framework. Results indicate that the DCPA and HRNetV2 combination surpasses DeepLabV3 and HRNetV2 models for active fire detection. In addition, the automated framework can deliver active fire detection results of Sentinel-2 inputs with coverage of about 12,000 km^2 (including data download) in less than 6 min, where average intersections over union (IoUs) of 70.4% and 71.9% were achieved in tests over Australia and the United States, respectively. Concepts in this framework can be further applied to other remote sensing sensors with data acquisitions in SWIR-NIR-Red ranges and can serve as a powerful tool to deal with large volumes of high-resolution data used in future fire monitoring systems and as a cost-efficient resource in support of governments and fire service agencies that need timely, optimized firefighting plans.

Keywords: active fire detection; Sentinel-2; multi-spectral; deep learning; dataset

1. Introduction

As a critical global phenomenon, wildfires are being exacerbated by the influence of climate change and are causing large amounts of economic and environmental damage each year [1]. For instance, Australia experienced one of its most devastating fire seasons in 2019–2020, colloquially known as "Black Summer" [2]. Commencing in June 2019, the fires were out of control across Australia from September 2019 to March 2020 causing at least 33 deaths, a total burned area of almost 19 million hectares, and economic loss of over 103 billion AUD [3]. Likewise, the United States has also experienced a series of wildfires since May 2020. By the end of that year, at least 46 deaths, a total burned area of over 4 million hectares, and an economic cost of almost 20 billion USD have been reported [4]. This dramatic increase in the number and severity of wildfire events around the world is promoted by the reduced live fuel moisture content [5], soil moisture [6], precipitation [7], and so forth, urging more than ever before the governments and fire management agencies to develop a reliable, timely, and cost-efficient fire monitoring system [8,9] providing

information on the status of wildfires globally. Associated Earth observation satellites have been deployed to achieve this in two aspects: burned area mapping and active fire detection [10–13]. This study will focus on the second aspect and detect the location of actively burning spots during the wildfire. Most prior studies on active fire detection take advantage of thermal infrared and mid-infrared bands with coarse spatial resolutions acquired by sensors like the advanced very high resolution radiometer (AVHRR), moderate resolution imaging spectroradiometer (MODIS), visible infrared imaging radiometer suite (VIIRS), sea and land surface temperature radiometer (SLSTR), and visible and infrared radiometer (VIRR) onboard near-polar orbit satellites or the advanced baseline imager (ABI), spinning enhanced visible and infrared imager (SEVIRI), and space environment data acquisition monitor (SEDI) onboard geostationary satellites [14–21]. These data provide consistent global monitoring in near-real-time with coarse spatial resolutions and are employed in several global monitoring systems like the Digital Earth Australia Hotspots from Geoscience Australia (GA), the Fire Information for Resource Management System (FIRMS) from the National Aeronautics and Space Administration (NASA), and the European Forest Fire Information System (EFFIS) from the European Space Agency (ESA). However, the background radiance in coarse resolution pixels tends to overwhelm signals from subtle hotspots due to the large instantaneous field of view (IFOV), and hence it is difficult to reveal active fires with more spatial details [22]. Further efforts have been placed on active fire detection using medium resolution acquisitions from Landsat and/or Sentinel-2 to increase spatial details and several algorithms have been developed based on reflective wavelength bands [11,12,22,23]. The Landsat-8 Operational Land Imager (OLI) data are mostly used to detect active fires at a spatial resolution of 30 m. The latest GOLI algorithm [12] is proposed based on a statistical examination on the top of atmosphere (TOA) reflectance of the red and SWIR bands and exhibited comparable commission error (CE) and slightly lower omission error (OE) over visual interpretations of detected active fire locations than other algorithms. This statistics-based method has been further applied to the Sentinel-2 acquisitions and an active fire detection method based on Sentinel-2 (AFD-S2) [23] is proposed for active fire detection at a spatial resolution of 20 m.

Although active fire detection using satellite imagery has been well established over the years, most detection algorithms mentioned above are based on supervised pixel- and/or region-level comparison involving sensor-specific thresholds and neighbourhood statistics [24]. With the increasing number of satellites launched and the volume of remote sensing data acquired, these methods will bring massive workloads for global or regional wildfire monitoring. Fast-growing deep-learning techniques [25] provide great tools for large-volume remote sensing data processing, inspiring us to explore their potential in active fire detection. This is a relatively new field and still lacks large-scale datasets and architectures for evaluation. The first contribution to this field is based on the Landsat-8 imagery [24], which introduces a large-scale dataset to investigate how deep-learning-based segmentation models with U-Net architecture [26] can be used to approximate the handcrafted active fire maps. Our study will focus on investigating how deep learning models can be applied to Sentinel-2 imagery to approximate the visual-interpretation-based active fire detection by formulating it as a binary semantic segmentation task. A dual-domain channel-position attention (DCPA) network is specifically designed based on the unique characteristics of active fires in false-colour composite images to further boost the detection accuracy. A dataset is constructed using the open-access Sentinel-2 product collected over wildfires on the east coast of Australia and the west coast of the United States in 2019 and 2020 for training and testing. Based on the developed best-performing active fire detection model, an automated active fire detection framework with three basic parts (data collection and preprocessing, deep-learning-based active fire detection, and final product generation) is proposed. This framework can automatedly output active fire detection results in Geotiff format using the geographic location and time inputs, and this Geographic Information System (GIS)-preferred format can be easily visualized and post-processed in software like ArcGIS [27], QGIS [28], etc.

2. Test Sites and Datasets

The Sentinel-2 constellation is part of the ESA's Copernicus Program, providing high-quality multi-spectral Earth observations since 2015. The twin satellites share a sun-synchronous orbit at a 786 km altitude with a repeat cycle of 10 days for one satellite and 5 days for two [29]. Each satellite carries a multi-spectral instrument (MSI) with 13 bands in the visible, near-infrared (NIR), and short-wave infrared (SWIR) ranges [30]. Prior studies have evidenced that active fires with larger sizes and higher temperatures normally exhibit higher top of atmosphere (TOA) reflectances in SWIR bands with negligible contributions in the red band [12,23]. Therefore, active fire detection using Sentinel-2 data can be realized by filtering out pixels with high values in B12 and B11 bands and low values in B4 as in the AFD-S2 method. Details of Sentinel-2/MSI bands involved in this study are listed in Table 1, where all four bands are involved in the generation of ground truth annotations during the dataset construction; B12, B11, and B8A are employed to generate the special indices for sifting background patches in the pre-processing stage (Section 3.1); and B12-B11-B4 are used as inputs of deep-learning models in the training and testing phases of the active fire detection (Section 3.2).

Table 1. List of Sentinel-2/MSI band information used in this study.

Channel	Sentinel-2A		Sentinel-2B		Spatial Resolution
	Central Wavelength	Bandwidth	Central Wavelength	Bandwidth	
4—Red	664.6 nm	31 nm	664.9 nm	31 nm	10 m
8A—Narrow NIR	864.7 nm	21 nm	864.0 nm	22 nm	20 m
11—SWIR 1	1613.7 nm	91 nm	1610.4 nm	94 nm	20 m
12—SWIR 2	2202.4 nm	175 nm	2185.7 nm	185 nm	20 m

A total of 135 Sentinel-2 Military Grid Reference System (MGRS) tiles over the states of New South Wales (NSW), Australia Capital Territory (ACT), and Victoria (VIC) covering areas of about 378,543 km^2 from 18 October 2019 to 14 January 2020, and 178 Sentinel-2 MGRS tiles over the state of California covering areas of about 457,594 km^2 from 28 July 2020 to 12 October 2020, were used in the construction of the dataset. Figure 1a,b illustrate Sentinel-2 coverage in Australia and the United States, respectively. The downloaded Sentinel visible bands have spatial resolutions of 10 m and associated NIR and SWIR bands have spatial resolutions of 20 m. Hence, visible bands were down-sampled to the same spatial resolution as the NIR and SWIR bands before dataset construction. Related ground truth annotations were generated by the synergy of the AFD-S2 method and visual inspection of Sentinel-2 false-colour composites. Visual inspection has been widely used in prior studies to generate ground truth annotations [11,12,23,24] because actively burning fires are ephemeral and sparsely located over a large area and therefore it is not practical to collect a large number of ground-based measurements at the time of satellite overpass.

In AFD-S2, a single criterion in (1) was applied for the active fire detection in temperate conifer forests on the east coast of Australia, whereas multiple criteria in (2) were used in Mediterranean forests, woodlands, and scrubs on the west coast of the United States.

$$0.504 * \rho_{SWIR_2} - \rho_{Red} \geq a \tag{1}$$

$$0.734 * \rho_{SWIR_2} - \rho_{Red} \geq b, \ \rho_{SWIR_2} \geq c, \ \rho_{SWIR_1} \geq d \ or \ \rho_{SWIR_2} \geq 1.0 \tag{2}$$

An important part is to determine the correct threshold to differentiate the active fire and background pixels. In this study, a, b, c, and d were initialized as 0.198, 0.068, 0.355, and 0.475 as mentioned in [23], which were further fine-tuned referring to the visual interpretation of B12-B11-B4 and B12-B11-B8A false-colour composites. Manual corrections were also performed on the obvious mislabelling to reduce OEs and CEs in the generated annotation map.

Finally, a dataset consisting of false-colour patches and associated dense pixel annotations of 2 classes (active fire and background) was constructed by clipping and sifting the Sentinel red-SWIR composites and the associated annotation map above.


(**a**)   (**b**)
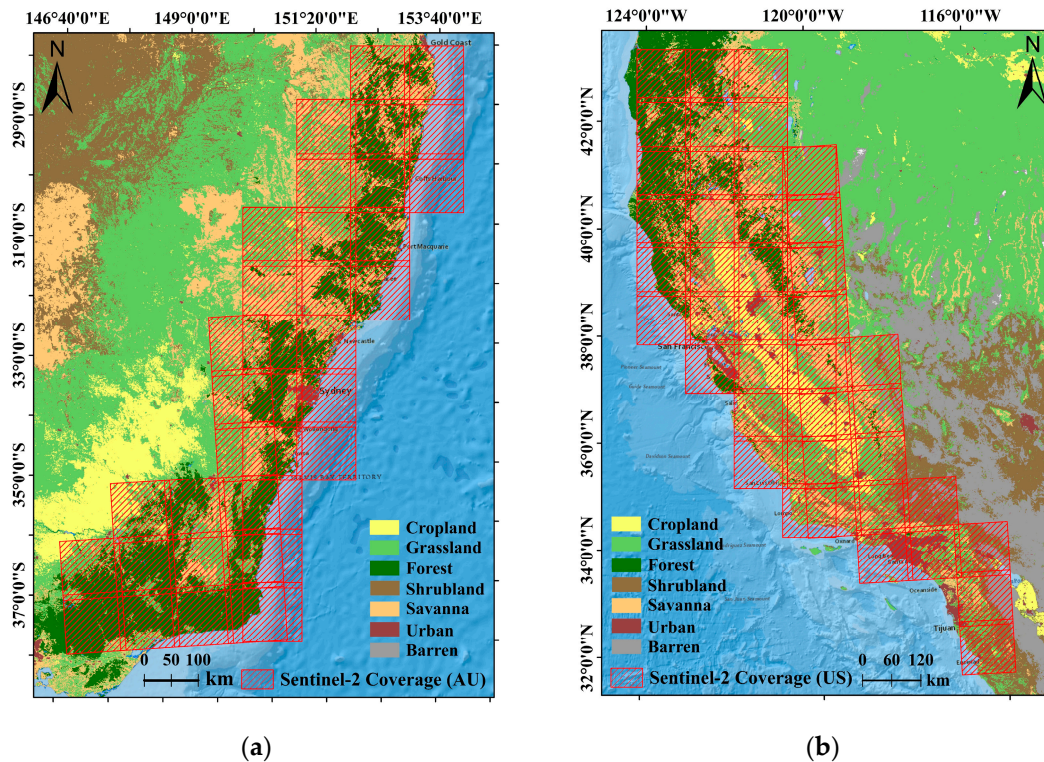
Figure 1. Sentinel-2 image coverage used for constructing dataset: (**a**) Australia; (**b**) the United States. Base map: land cover map derived from the 2019 Version 6 MODIS Land Cover Type (MCD12Q1) product [31].

3. Methodology

3.1. Automated Active Fire Detection Framework

Details of the proposed automated active fire detection framework are shown in Figure 2, along with data collection and preprocessing, deep-learning-based active fire detection, and product generation modules included in yellow, green, and blue boxes, respectively.
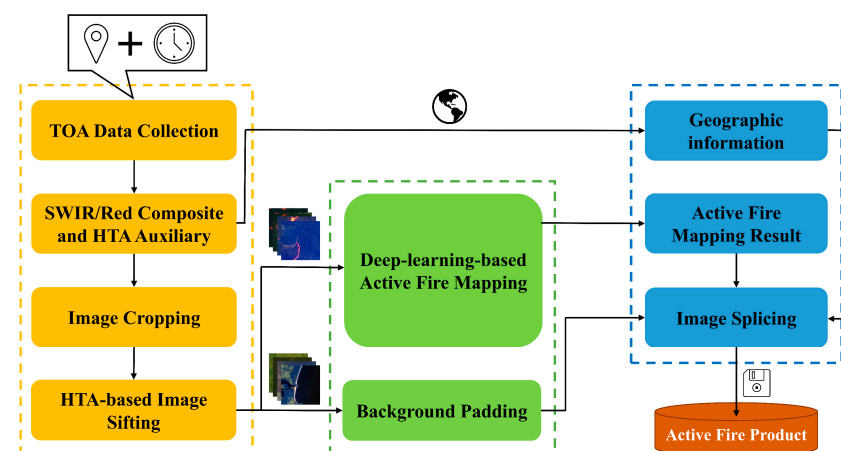
Figure 2. Illustration of the automated active fire detection framework.

In the yellow box, Sentinel-2 TOA reflectance products can be automatically downloaded using the time and location information provided by users through the API offered

by ESA. SWIR and red bands are composited into false-colour images with salient active fires and the normalized hotspot index (NHI) images are generated based on (3) and (4) [32,33] (alterations of normalized burned ratios (NBR)) as auxiliaries in the upcoming sifting step to coarsely separate between background paddings and patches that might have high-temperature anomalies (HTAs) like wildfires, industrial heat sources, volcanic activities, land management fires, etc. Criteria for HTA detection here are slightly looser than those in [32,33] to make sure cool smouldering fires are included in the generated HTA maps. These HTA maps are further fed into the image cropping step along with the false-colour combinations.

$$NHI_{SWIR} = \frac{SWIR_2 - SWIR_1}{SWIR_2 + SWIR_1} \tag{3}$$

$$NHI_{SWNIR} = \frac{SWIR_1 - NIR}{SWIR_1 + NIR} \tag{4}$$

All images are cropped into small patches with identical sizes and sequential identities, which are further allocated to the active fire and background datasets, respectively. Whether a false-colour patch is active or not is determined by checking if there is a positive value in the associated HTA patch. This dataset splitting strategy avoids redundant processing in the background area and hence reduces the time and memory cost of the framework. Besides that, the synergy of image cropping and dataset splitting can effectively reduce the imbalance between the active fire and background proportions in each false-colour patch. In subsequent steps, active false-colour patches are fed into the trained deep-learning-based models to generate the segmentation results while the background ones only deliver their identities to the padding images for further splicing. The segmentation results and associated background paddings are further projected back into the original frame to generate final products in Geotiff format, where geographical information extracted from the original data are added into these spliced products.

3.2. Deep-Learning-Based Active Fire Detection

Active fire pixels normally appear in bright orange in a SWIR-red false-colour image because of the higher reflectance in SWIR bands and the lower reflectance in the red band [11,34,35]. After igniting, fires quickly spread around and turn into local clusters of different sizes. The sharp contrast between active fire and background allows the detection task to be considered as a binary semantic segmentation of the false-colour image, even though special attention needs to be paid to the unique behaviours of the active fire. For example, OEs are usually caused by small and cool fires, CEs are likely to occur in soil-dominated pixels or highly reflective building rooftops, and proportions of the active fire and background pixels are unbalanced [36].

We first used two deep-learning-based semantic segmentation models with different architectures to detect the active fire from the false-colour composite image. One of them is the DeepLabV3+ model with an encoder-decoder architecture [37–39], where the encoder module encodes multi-scale contextual information by applying parallel atrous convolutions with different rates while the decoder recovers high-resolution representations and refines the segmentation result along the object boundaries. Though pyramid features are fused by the atrous spatial pyramid pooling (ASPP) module in DeepLabV3+, there are still losses of spatial information during the feature embedding, which causes challenges in the reconstruction of segmentation results with original sizes. The other model we used for active fire detection is HRNetV2, which maintains the high-resolution representations throughout the whole architecture by parallelly connecting high-to-low resolution convolution streams and repeatedly exchanging information across different resolutions [40,41]. Compared with DeepLabV3+, representations in HRNetV2 have richer semantic and spatial information resulting in more precise boundaries in the segmentation results. To make HRNetV2 more applicable to the active fire detection task and better balance OEs and CEs in detection results, a DCPA network is further proposed by connecting a dual-domain

feature extractor to the existing HRNetV2 model using a channel-position attention block as presented in Figure 3.
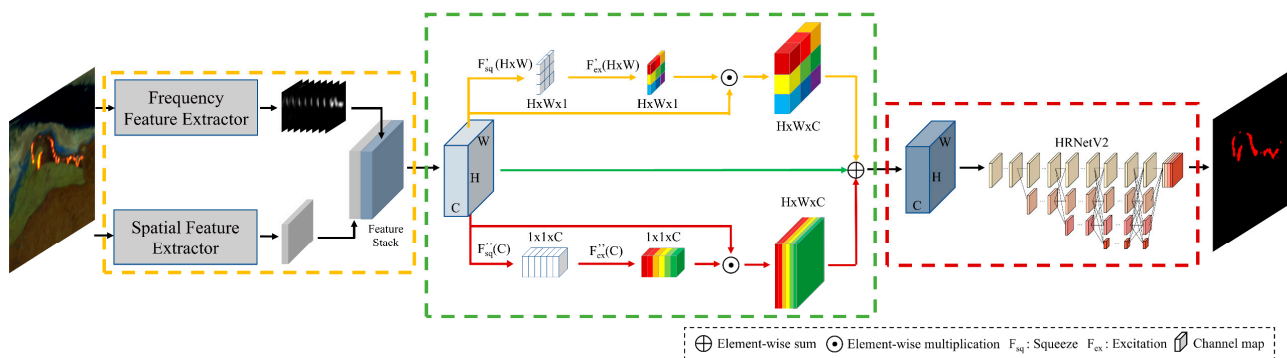
Figure 3. The architecture of the proposed DCPA+HRNetV2 active fire detection model.

The proposal of the DCPA network was inspired by the principle of visual inspection. Modifications mainly focus on the first few layers of the deep neural network that is normally responsible for the extraction of simple features. During active fire detection, human attention is automatically placed on the salient regions that present a sharp contrast with the background. In addition, boundaries of the active fires inside are outlined using spatial distributions and interdependences of the associated fire pixels. Attention on the salient regions helps to locate anomalous regions and hence reduce the OEs in the sparse and cool fires, while statistics in the spatial distribution and interdependence offer a better distinction between active fires and surrounding pixels, which helps to reduce the CEs caused by the ambiguity between hot soil and cooler active fire.

Therefore, a dual-domain feature extractor is first constructed in the DCPA network as shown in the yellow box of Figure 3. The spatial feature extractor is constructed based on the deep convolutional neural network (DCNN) [42], where a 3×3 convolution layer, a 1×1 convolution layer, and a Resblock [43] are assembled to extract the spatial distribution and interdependence among pixels. The frequency feature extractor is proposed based on the visual saliency paradigm in [44], which mainly focuses on generating saliency maps by smoothing the amplitude spectrum derived from the hypercomplex Fourier transform (HFT). To construct the frequency feature extractor, we replace the imaginary parts in the quaternion matrix f [44] with the reflectance of red and SWIR bands.

$$f(n,m) = \omega_0 P_0 + \omega_1 P_1 i + \omega_2 P_2 j + \omega_3 P_3 k \tag{5}$$

$$P_1 = \rho_{SWIR_2},\ P_2 = \rho_{SWIR_1},\ P_3 = \rho_{Red} \tag{6}$$

where ω_0, ω_1, ω_2, and ω_3 are the weights of the feature matrices P_0, P_1, P_2, and P_3, and i, j, k are imaginary units in quaternion form ($i^2 = j^2 = k^2 = ijk = -1$, $ij = k$, $ik = j$, $jk = i$). The motion feature P_0 of a static input is equal to zero and makes no contribution to the quaternion matrix f. Weights of spectral feature matrices are set as $\omega_1 = 0.2$, $\omega_2 = 0.4$, and $\omega_3 = 0.4$. The HFT of the quaternion matrix f can be generated through

$$F_H[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu 2\pi(mv/M + nu/N)} f(n,m) = A(u,v)e^{\chi\Phi(u,v)} \tag{7}$$

where $\mu = (i + j + k)/\sqrt{3}$ is a unit pure quaternion, $A(u,v)$ and $\Phi(u,v)$ are the amplitude and phase spectrums of $F_H[u,v]$, and χ is a pure quaternion matrix as the eigenaxis spectrum [45]. It has been proven that texture-rich salient regions like active fire are more likely to be embedded in the background of the amplitude spectrum whereas the repeated and uniform patterns such as the spatial background are prone to appear as sharp spikes [46]. Therefore, frequency features of different salient regions can be effectively

enhanced by smoothing the amplitude spectrum with a series of Gaussian kernels $\{g_k\}$ at different scales [44].

$$g_k(u,v) = \frac{1}{\sqrt{2\pi}2^{k-1}\tau_0} e^{-(u^2+v^2)/(2^{2k-1}\tau_0^2)} \tag{8}$$

$$\Lambda_k(u,v) = (g_k(u,v) \star A)(u,v) \tag{9}$$

where $\tau_0 = 0.5$, $k = 1, \ldots, \log_2 min\{H, W\} + 1$, and H and W are the height and width of the image. Then, the saliency maps $\{s_k \in \mathbb{R}^{H \times W}\}$ can be generated by the inversed HFT.

$$s_k = \left\| \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{\mu 2\pi(mv/M + nu/N)} \Lambda_k(u,v) e^{\chi \Phi(u,v)} \right\|^2 \tag{10}$$

Outputs of the dual-domain feature extractor constitute a feature stack $S \in \mathbb{R}^{H \times W \times C}$ with a channel depth of C, which is further recalibrated by the channel-position attention as shown in the green box of Figure 3. The channel-position attention consists of a channel-wise and a position-wise squeeze-and-excitation (SE) block [47], which explicitly model interdependencies between channels or positions and use the global information to selectively emphasize informative features and suppress less useful ones. In the yellow stream, the global channel information is first squeezed into a position descriptor by average pooling in the channel dimension.

$$x = F'_{sq}(S) = \frac{1}{C} \sum_{k=1}^{C} S(:,:,k) \tag{11}$$

An excitation operation is further performed on the squeezed output x to capture the position-wise dependencies. This is achieved by a simple gating mechanism with the ReLU activation function $\delta(\cdot)$.

$$x_1 = F'_{ex}(x) = \delta(W_1(\delta(W_2 x))) \tag{12}$$

where $W_1 \in \mathbb{R}^{H \times W \times \frac{H \times W}{d}}$ and $W_2 \in \mathbb{R}^{\frac{H \times W}{d} \times H \times W}$ are the weights of the two convolution layers with d set as 256 for the best performance. The output of the yellow stream S' is then obtained by rescaling S with the excitation output x_1.

In the red stream, the average pooling is initially applied in the spatial dimension to obtain the squeezed channel descriptor. This is followed by an excitation operation to obtain the channel-wise dependencies.

$$y = F''_{sq}(S) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} S(i,j,:) \tag{13}$$

$$y_1 = F''_{ex}(y) = \sigma(W_3(\delta(W_4 y))) \tag{14}$$

where, $\sigma(\cdot)$ is the sigmoid activation function, and $W_3 \in \mathbb{R}^{C \times \frac{C}{r}}$ and $W_4 \in \mathbb{R}^{\frac{C}{r} \times C}$ are the weights of the two fully connected layers with $r = 16$ for the best performance. Likewise, the output of this stream S'' is obtained by rescaling S with the excitation output y_1. The original feature stack S is maintained through the green stream in the fusion process to tackle the gradient vanishing issue and to boost feature discriminability. The outputs of these three streams are then weighted summed to generate the segmentation input.

$$\widetilde{S} = S + \alpha S' + \beta S'' \tag{15}$$

where α and β are initialized as 0 and gradually increase during the training process. The weighted feature $\widetilde{S}$ is then fed into the HRNetV2 model to obtain the active fire detection result.

3.3. Evaluation Metrics

Quantitative metrics commonly used in semantic segmentation and active fire detection such as CE, OE, and intersection over union (IoU) are calculated to evaluate the performance. All these metrics can be directly derived from the elements in the confusion matrix, which are the numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FP) pixels. In an active fire detection task, TP denotes the correctly labelled active fire pixel, FP is the background pixel incorrectly labelled as active fire, TN stands for the correctly labelled background pixel, and FN is the active fire pixel incorrectly labelled as background.

Based on Table 2, the IoU, CE, and OE can be calculated as

$$IoU = \frac{TP}{TP + FP + FN}, \quad CE = \frac{FP}{TP + FP}, \quad OE = \frac{FN}{TP + FN} \tag{16}$$

Table 2. Confusion matrix in active fire detection.

| Classified	Reference Active Fire	Background
Active Fire	TP	FP
Background	FN	TN

3.4. Implementation Details

The deep-learning-based models (DeepLabV3+, HRNetV2, DCPA+HRNetV2) for active fire detection were trained and tested on the constructed dataset. After clipping and sifting, this dataset had 37,016 samples with 21,342 of them from Australia (AU) and 15,674 of them from the United States (US). These patches were randomly divided into three subsets with 60% for training, 10% for validation, and the remaining 30% for testing. Each sample in the constructed dataset had a false-colour patch with the size of $128 \times 128 \times 3$ and a dense pixel annotation with the size of 128×128. During the training process, inputs were augmented by random flipping and rotating and an Adam optimizer [48] was initialized with a base learning rate of 0.01 and a weight decay of 0.0005. The poly learning rate policy with the power of 0.9 was used to drop the learning rate. All experiments were implemented in the PyTorch framework on two NVIDIA GeForce RTX 2080Ti GPUs with a batch size of 16.

4. Experiment Results

4.1. Deep-Learning-Based Active Fire Detection

To illustrate the performance of three deep-learning-based semantic segmentation models (DeepLabV3+, HRNet-V2, and DCPA+HRNetV2) on the active fire detection task, sample outputs of these three models are first shown in Figure 4. Related accuracy metrics were calculated over the testing dataset (see Section 3.4) and are listed in Table 3 along with the backbone used in each model. Referring to [38], Xception-71 with one entry flow, multiple middle flows, and one exit flow was used as the backbone in the DeepLabV3+ model, where all the max pooling operations in the original Xception [49] were replaced by a combination of depthwise separable convolution with striding, batch normalization [50], and ReLU [51]. Following [41], HRNetV2-W48 with four stages was used as the backbone in DCPA+HRNetV2 and HRNetV2. Each stage is composed of repeated modularized blocks in branches associated with different resolutions, where the multi-resolution features are fused by the fusion unit [40].
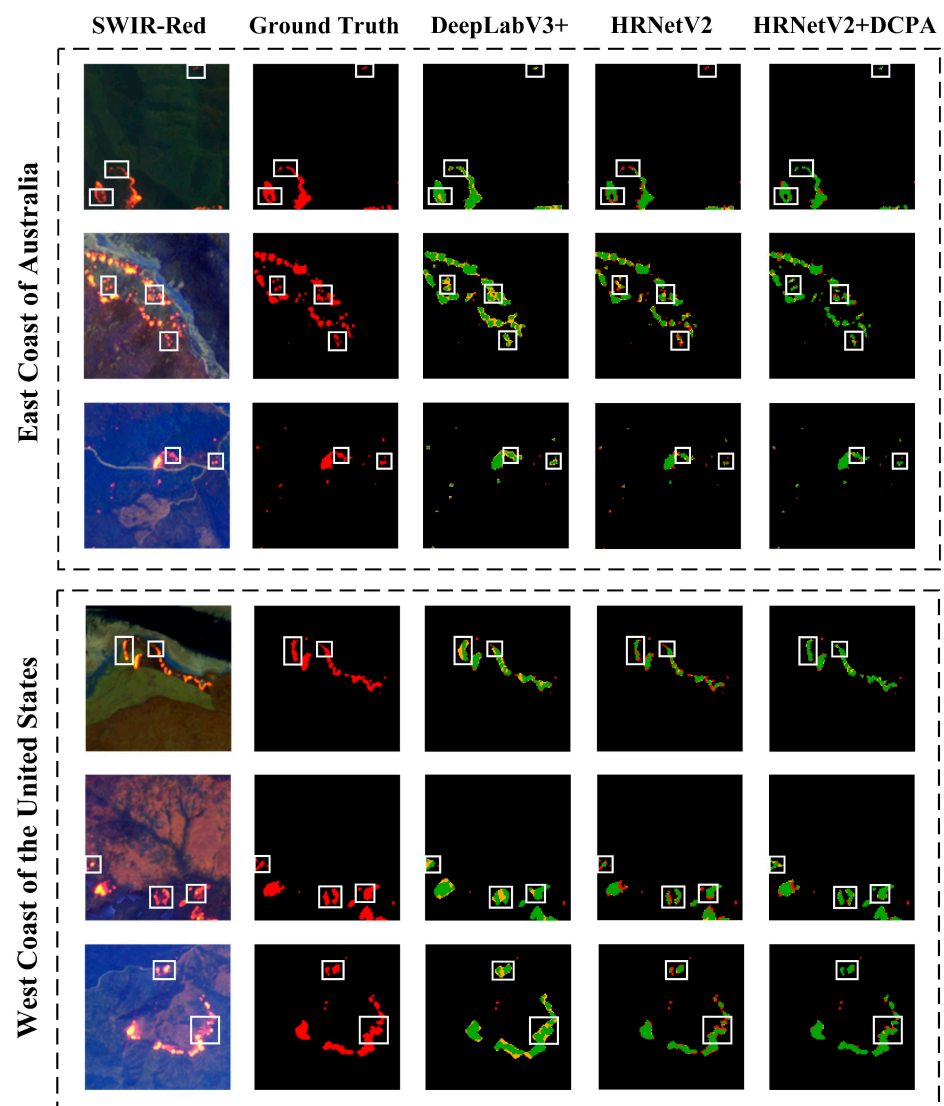
Figure 4. Active fire samples (false-colour composite and annotation) and the associated segmentation results derived from different models (green: TP; red: FN; yellow: FP).

Table 3. Accuracy metrics of the active fire detection results.

Model	Backbone	OE [a]	CE [a]	IoU [a]	OE [b]	CE [b]	IoU [b]
DeepLabV3+	Xception-71	7.6%	28.1%	67.8%	9.1%	25.3%	69.5%
HRNetV2	HRNetV2-W48	18.5%	14.2%	71.8%	19.0%	11.5%	73.2%
DCPA+HRNetV2	HRNetV2-W48	17.3%	13.1%	73.4%	17.4%	9.2%	76.2%

[a] East coast of Australia (AU); [b] west coast of the United States (US).

Active fire detection results in Figure 4 indicate that HRNetV2 models focus more on the spatial details in the segmentation result compared with the DeepLabV3+ by paying more attention to depicting boundaries and edges and preserving high spatial resolution representations in the propagation. For example, as presented in the white boxes of Figure 4, background areas between adjacent small fires are likely to be mislabelled as fire areas in DeepLabV3+ results, because details of these adjacent small fires are merged during the semantic information extraction and are not properly recovered through the decoder. This inevitably increases the CE in DeepLabV3+ results as indicated in Table 3. Though the CE in HRNetV2 results is much lower than in DeepLabV3+, a higher OE can be observed because too much attention has been paid to describing the boundaries of high-temperature active fires and thus the small and cooler fires have been neglected. As

shown in the first row of Figure 4, the isolated small fire has been detected by DeepLabV3+ but not by HRNetV2. This is the main reason for introducing the DCPA network, which can focus on the salient area and meanwhile properly allocate attention to small or/and cooler fires during the segmentation. It is evidenced in both results in Figure 4 and Table 3 that the introduction of the DCPA network helps to locate sparse and isolated active fires with only a few pixels and balance the OEs and CEs for a better IoU.

4.2. Automated Active Fire Detection Framework

The best-performing DCPA+HRNetV2 model was assembled into the proposed framework for the automated active fire detection based on Sentinel-2 imagery. Results were automatically generated in Geotiff format through inputting locations of the bounding box and the time of the fire event. With the details listed in Table 4, tests were carried out on the east coast of Australia and the west coast of the United States. Final products were zoomed into areas with dense active fire pixels and laid over the false-colour base image in ArcMap as presented in Figure 5.

Table 4. Detection accuracy and processing efficiency of the automated framework.

East Coast of Australia				
Acquisition Date	2020-01-15	2020-01-25	2020-01-28	2020-01-30
Top Left [a]	−35.5, 148.7	−37.3, 148.5	−36.3, 147.6	−34.8, 148.7
Top Right [a]	−35.5, 150.2	−37.3, 149.2	−36.3, 148.7	−34.8, 150.6
Bottom Left [a]	−36.8, 148.7	−37.8, 148.5	−36.9, 147.6	−37.6, 148.7
Bottom Right [a]	−36.8, 150.2	−37.8, 149.2	−36.9, 148.7	−37.6, 150.6
IoU	71.4%	71.1%	70.2%	68.7%
Time Cost	358 s	105 s	133 s	839 s
West Coast of the United States				
Acquisition Date	2020-10-13	2020-10-18	2020-10-28	2020-10-29
Top Left [a]	37.2, −119.4	37.6, −118.9	37.6, −119.4	41.9, −123.7
Top Right [a]	37.2, −118.6	37.6, −118.2	37.6, −118.9	41.9, −123.3
Bottom Left [a]	36.7, −119.4	37.3, −118.9	37.3, −119.4	41.0, −123.7
Bottom Right [a]	36.7, −118.6	37.3, −118.2	37.6, −118.9	41.0, −123.3
IoU	72.6%	71.3%	73.0%	70.8%
Time Cost	58 s	51 s	46 s	132 s

[a] Location of the corner pixels (latitude, longitude) in UTM projection and WGS84 geodetic system.

Evaluation of the proposed automated active fire detection framework was carried out concerning detection accuracy and processing efficiency. As indicated in Table 4, the automated active fire detection framework achieved average IoUs of 70.4% over Australia and 71.9% over the United States. The most time-consuming parts of the framework are the data collection and active fire detection steps. For a Sentinel-2 input with total coverage of around 12,000 km^2, it takes about 350 s to obtain the final output with the active fire detection process taking up 140 s. This time-consuming detection process underlines the necessity of the active fire and background dataset separation in the preprocessing module. All the operations above were undertaken in configurations of an Intel Core i7-9700k processer with a base frequency of 3.60 GHZ, NVIDIA GeForce RTX 2080Ti GPU, and a system memory of 64 GB.
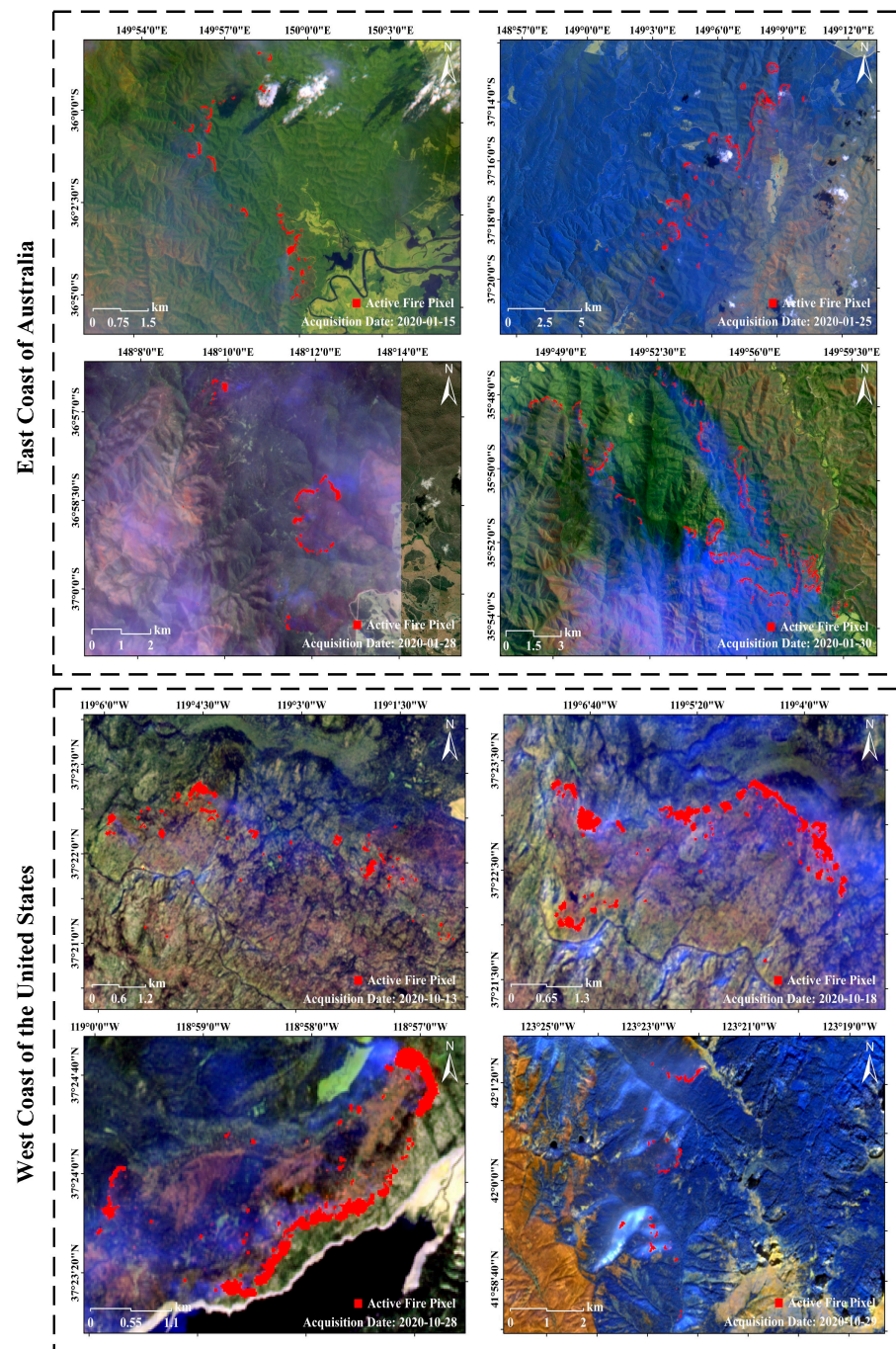
Figure 5. Test results of the automated framework on different dates over Australia and the United States.

5. Discussion

The accuracy of the active fire product generated by the proposed automated framework can be affected by several factors including the segmentation algorithm, annotation method, data preprocessing step, data source, etc. First of all, compared with semantic segmentation applications in other fields, active fire detection can be simplified as a binary segmentation task, although it is of increased difficulty as it deals with a large number of small fire objects. These small fires will increase the possibility of OE in the result. In addition, labels in the constructed dataset are derived from the synergy of the AFD-S2 method and visual inspection of Sentinel-2 band combinations. Although the time and labour involved are significantly reduced by this synergy, there are still some omitted

small and cool fires with lower radiances or mislabelled soil-dominated pixels and highly reflective building rooftops derived from the AFD-S2 method that are not corrected during the visual inspection. This may cause deep-learning-based models to learn incorrect semantic patterns during active fire detection and thus lead to OE and CE in the result. Moreover, although criteria for the HTA detection are slightly loosened during the splitting of the active fire patch and background padding, there are still some cool smouldering fires that are not included in the generated HTA map, and hence misclassified into the background dataset. Another source of error is intrinsic in the SWIR/Red-based active fire detection, which mainly uses the sharp contrast between high-temperature active fire and background in TOA bands. However, the existence of heat radiation results in ambiguities between classes affecting the classification accuracy of pixels along edges, and consequently introduces errors in the segmentation results. By way of example, Figure 6 zooms over areas with small and cool fires and highly reflective building rooftops to present the OEs and CEs in the results.

(**a**)   (**b**)
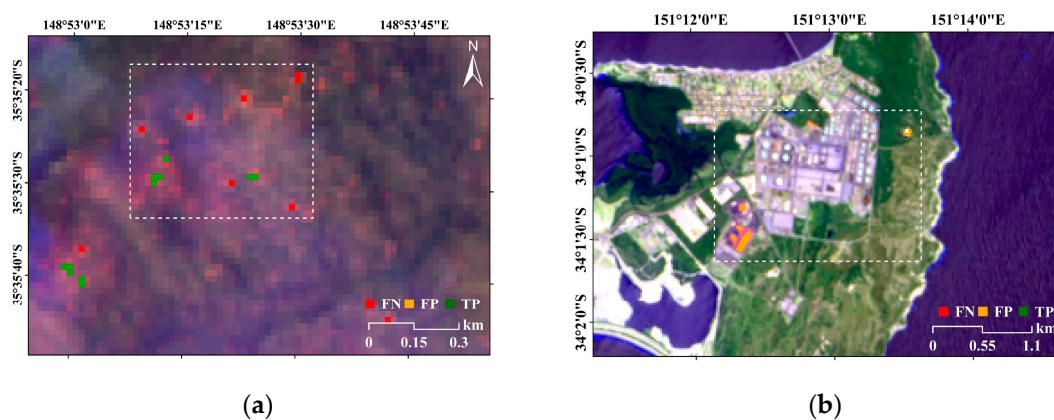
Figure 6. Examples of OEs and CEs in the active fire detection result: (**a**) OE; (**b**) CE.

Despite the aforementioned errors, the proposed framework outperforms the current fire monitoring systems described in Section 1 in terms of delivering spatial details, which can help to calibrate or/and improve the current low-resolution results. Figure 7 gives an example of our Sentinel-2 and the MODIS Collection 6.1 [17] active fire products. Associated Sentinel-2 and MODIS acquisitions are collected on the same day over the same area. It can be concluded that the MODIS active fire product can only provide a coarse depiction of the active fire because of the low spatial resolution of 1 km. Our Sentinel-2 product, however, offers more detail (location, boundary, and shape, etc.) on active fires even for small fires.

(**a**)   (**b**)
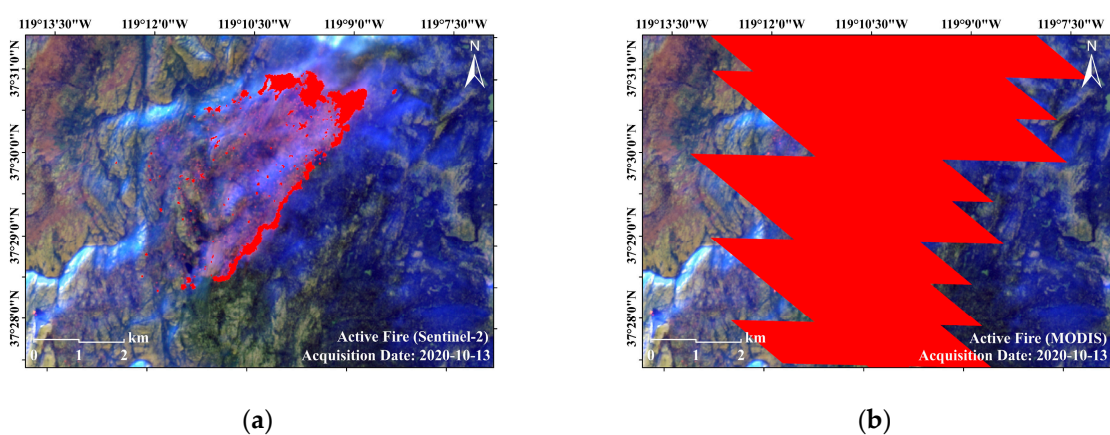
Figure 7. Example of active fire product: (**a**) Sentinel-2 product generated by the proposed framework; (**b**) MODIS Collection 6.1.

6. Conclusions

An automated active fire detection framework with data collection and preprocessing, deep-learning-based active fire detection, and final product generation modules was developed in this research serving as a prototypical function for a future fire monitoring system. The active fire detection module in the framework was developed on a specifically designed DCPA+HRNetV2 network, which is trained on the dataset constructed using SWIR, NIR, and red bands in Sentinel-2 Level-2C products. Results indicate that this DCPA and HRNetV2 combination outperformed DeepLabV3 and HRNetV2 models on the active fire detection, and the automated framework can effectively deliver active fire detection results with an average IoU higher than 70%. Although in the current state the refresh rate of this framework is only 5 days and it is only suitable for inputs of Sentinel-2 acquisitions over the east coast of Australia and the west coast of the United States, the spatial resolution (20 m) of the outputs is much higher than the results derived from current fire systems. With the launch of more and more high-resolution and super-spectral sensors onboard remote sensing satellites, high-quality and near-real-time Earth observations will be achieved in the future. At that time, an automated framework like this will offer a powerful tool and a cost-efficient resource in support of the governments and fire service agencies that need timely, optimized firefighting plans.

References

1. Artés, T.; Oom, D.; De Rigo, D.; Durrant, T.H.; Maianti, P.; Libertà, G.; San-Miguel-Ayanz, J. A global wildfire dataset for the analysis of fire regimes and fire behaviour. *Sci. Data* **2019**, *6*, 296. [CrossRef] [PubMed]
2. Kemter, M.; Fischer, M.; Luna, L.; Schönfeldt, E.; Vogel, J.; Banerjee, A.; Korup, O.; Thonicke, K. Cascading hazards in the aftermath of Australia's 2019/2020 Black Summer wildfires. *Earth's Future* **2021**, *9*, e2020EF001884. [CrossRef]
3. Killian, D. Black Summer in the Shoalhaven, NSW: What was it like and what could we do better? A resident's perspective. *Public Health Res. Pract.* **2020**, *30*, e3042028. [CrossRef]
4. Higuera, P.E.; Abatzoglou, J.T. Record-setting climate enabled the extraordinary 2020 fire season in the western United States. *Glob. Chang. Biol.* **2021**, *27*, 1–2. [CrossRef] [PubMed]
5. Ruffault, J.; Martin-StPaul, N.; Pimont, F.; Dupuy, J.-L. How well do meteorological drought indices predict live fuel moisture content (LFMC)? An assessment for wildfire research and operations in Mediterranean ecosystems. *Agric. For. Meteorol.* **2018**, *262*, 391–401. [CrossRef]
6. Sungmin, O.; Hou, X.; Orth, R. Observational evidence of wildfire-promoting soil moisture anomalies. *Sci. Rep.* **2020**, *10*, 11008.
7. Holden, Z.A.; Swanson, A.; Luce, C.H.; Jolly, W.M.; Maneta, M.; Oyler, J.W.; Warren, D.A.; Parsons, R.; Affleck, D. Decreasing fire season precipitation increased recent western US forest wildfire activity. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E8349–E8357. [CrossRef]
8. Hua, L.; Shao, G. The progress of operational forest fire monitoring with infrared remote sensing. *J. For. Res.* **2017**, *28*, 215–229. [CrossRef]

9. Jones, M.W.; Smith, A.; Betts, R.; Canadell, J.G.; Prentice, I.C.; Le Quéré, C. Climate change increases risk of wildfires. *ScienceBrief Rev.* **2020**, *116*, 117.

10. Giglio, L.; Csiszar, I.; Restás, Á.; Morisette, J.T.; Schroeder, W.; Morton, D.; Justice, C.O. Active fire detection and characterization with the advanced spaceborne thermal emission and reflection radiometer (ASTER). *Remote Sens. Environ.* **2008**, *112*, 3055–3063. [CrossRef]

11. Schroeder, W.; Oliva, P.; Giglio, L.; Quayle, B.; Lorenz, E.; Morelli, F. Active fire detection using Landsat-8/OLI data. *Remote Sens. Environ.* **2016**, *185*, 210–220. [CrossRef]

12. Kumar, S.S.; Roy, D.P. Global operational land imager Landsat-8 reflectance-based active fire detection algorithm. *Int. J. Digit. Earth* **2018**, *11*, 154–178. [CrossRef]

13. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. [CrossRef]

14. Stroppiana, D.; Pinnock, S.; Gregoire, J.-M. The global fire product: Daily fire occurrence from April 1992 to December 1993 derived from NOAA AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1279–1288. [CrossRef]

15. Wooster, M.J.; Xu, W.; Nightingale, T. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. *Remote Sens. Environ.* **2012**, *120*, 236–254. [CrossRef]

16. Schroeder, W.; Oliva, P.; Giglio, L.; Csiszar, I.A. The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sens. Environ.* **2014**, *143*, 85–96. [CrossRef]

17. Giglio, L.; Schroeder, W.; Justice, C.O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **2016**, *178*, 31–41. [CrossRef] [PubMed]

18. Lin, Z.; Chen, F.; Niu, Z.; Li, B.; Yu, B.; Jia, H.; Zhang, M. An active fire detection algorithm based on multi-temporal FengYun-3C VIRR data. *Remote Sens. Environ.* **2018**, *211*, 376–387. [CrossRef]

19. Roberts, G.; Wooster, M. Development of a multi-temporal Kalman filter approach to geostationary active fire detection & fire radiative power (FRP) estimation. *Remote Sens. Environ.* **2014**, *152*, 392–412.

20. Hall, J.V.; Zhang, R.; Schroeder, W.; Huang, C.; Giglio, L. Validation of GOES-16 ABI and MSG SEVIRI active fire products. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *83*, 101928. [CrossRef]

21. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn. Ser. II* **2016**, *94*, 151–183. [CrossRef]

22. Murphy, S.W.; de Souza Filho, C.R.; Wright, R.; Sabatino, G.; Pabon, R.C. HOTMAP: Global hot target detection at moderate spatial resolution. *Remote Sens. Environ.* **2016**, *177*, 78–88. [CrossRef]

23. Hu, X.; Ban, Y.; Nascetti, A. Sentinel-2 MSI data for active fire detection in major fire-prone biomes: A multi-criteria approach. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *101*, 102347. [CrossRef]

24. De Almeida Pereira, G.H.; Fusioka, A.M.; Nassu, B.T.; Minetto, R. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 171–186. [CrossRef]

25. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

27. Ormsby, T.; Napoleon, E.; Burke, R.; Groessl, C.; Feaster, L. *Getting to Know ArcGIS Desktop: Basics of ArcView, ArcEditor, and ArcInfo*; ESRI, Inc.: West Redlands, CA, USA, 2004.

28. *Qgis: A Free and Open Source Geographic Information System, Version 3.8.0*; Computer Software; QGIS Development Team: Zurich, Switzerland, 2018.

29. Bouzinac, C.; Lafrance, B.; Pessiot, L.; Touli, D.; Jung, M.; Massera, S.; Neveu-VanMalle, M.; Espesset, A.; Francesconi, B.; Clerc, S. Sentinel-2 level-1 calibration and validation status from the mission performance centre. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4347–4350.

30. Spoto, F.; Sy, O.; Laberinti, P.; Martimort, P.; Fernandez, V.; Colin, O.; Hoersch, B.; Meygret, A. Overview of Sentinel-2. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 1707–1710.

31. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [CrossRef]

32. Genzano, N.; Pergola, N.; Marchese, F. A Google Earth Engine tool to investigate, map and monitor volcanic thermal anomalies at global scale by means of mid-high spatial resolution satellite data. *Remote Sens.* **2020**, *12*, 3232. [CrossRef]

33. Marchese, F.; Genzano, N.; Neri, M.; Falconieri, A.; Mazzeo, G.; Pergola, N. A Multi-channel algorithm for mapping volcanic thermal anomalies by means of Sentinel-2 MSI and Landsat-8 OLI data. *Remote Sens.* **2019**, *11*, 2876. [CrossRef]

34. Miller, J.D.; Thode, A.E. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sens. Environ.* **2007**, *109*, 66–80. [CrossRef]

35. Gargiulo, M.; Dell'Aglio, D.A.G.; Iodice, A.; Riccio, D.; Ruello, G. A CNN-based super-resolution technique for active fire detection on Sentinel-2 data. In Proceedings of the 2019 PhotonIcs & Electromagnetics Research Symposium-Spring (PIERS-Spring), Rome, Italy, 17–20 June 2019; pp. 418–426.
36. Palacios, A.; Muñoz, M.; Darbra, R.; Casal, J. Thermal radiation from vertical jet fires. *Fire Saf. J.* **2012**, *51*, 93–101. [CrossRef]
37. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
38. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
39. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
40. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 5693–5703.
41. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [CrossRef] [PubMed]
42. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
44. Li, J.; Levine, M.D.; An, X.; Xu, X.; He, H. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 996–1010. [CrossRef]
45. Ell, T.A.; Sangwine, S.J. Hypercomplex Fourier transforms of color images. *IEEE Trans. Image Process.* **2006**, *16*, 22–35. [CrossRef]
46. Chen, B.; Shu, H.; Coatrieux, G.; Chen, G.; Sun, X.; Coatrieux, J.L. Color image analysis by quaternion-type moments. *J. Math. Imaging Vis.* **2015**, *51*, 124–144. [CrossRef]
47. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
48. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. *arXiv* **2019**, arXiv:1912.01703.
49. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
50. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
51. Laurent, T.; Brecht, J. The multilinear structure of ReLU networks. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 2908–2916.