






Article

# Performance Analysis of Deep Convolutional Autoencoders with Different Patch Sizes for Change Detection from Burnt Areas

Pablo Pozzobon de Bem <sup>1</sup>, Osmar Abílio de Carvalho Júnior <sup>1,\*</sup>,  
Osmar Luiz Ferreira de Carvalho <sup>2</sup>, Roberto Arnaldo Trancoso Gomes <sup>1</sup> and  
Renato Fontes Guimarães <sup>1</sup>

<sup>1</sup> Departamento de Geografia, Campus Universitário Darcy Ribeiro, Asa Norte, Universidade de Brasília, DF, Brasília 70910-900, Brazil; pablo.bem@aluno.unb.br (P.P.d.B.); robertogomes@unb.br (R.A.T.G.); renatofg@unb.br (R.F.G.)

<sup>2</sup> Departamento de Engenharia Elétrica, Campus Universitário Darcy Ribeiro, Asa Norte, Universidade de Brasília, DF, Brasília 70910-900, Brazil; osmarcarvalho@ieee.org

\* Correspondence: osmarjr@unb.br

Received: 14 July 2020; Accepted: 7 August 2020; Published: 11 August 2020



**Abstract:** Fire is one of the primary sources of damages to natural environments globally. Estimates show that approximately 4 million km<sup>2</sup> of land burns yearly. Studies have shown that such estimates often underestimate the real extent of burnt land, which highlights the need to find better, state-of-the-art methods to detect and classify these areas. This study aimed to analyze the use of deep convolutional Autoencoders in the classification of burnt areas, considering different sample patch sizes. A simple Autoencoder and the U-Net and ResUnet architectures were evaluated. We collected Landsat 8 OLI+ data from three scenes in four consecutive dates to detect the changes specifically in the form of burnt land. The data were sampled according to four different sampling strategies to evaluate possible performance changes related to sampling window sizes. The training stage used two scenes, while the validation stage used the remaining scene. The ground truth change mask was created using the Normalized Burn Ratio (*NBR*) spectral index through a thresholding approach. The classifications were evaluated according to the *F1* index, *Kappa* index, and mean Intersection over Union (*mIoU*) value. Results have shown that the U-Net and ResUnet architectures offered the best classifications with average *F1*, *Kappa*, and *mIoU* values of approximately 0.96, representing excellent classification results. We have also verified that a sampling window size of 256 by 256 pixels offered the best results.

**Keywords:** deep learning; CNN; classification; fire; multitemporal image

## 1. Introduction

Deep Learning (DL) is the term that refers to the use of multilayered neural networks to solve complex problems. It is one of the fastest-growing trends in machine learning, data science, and computer vision. The DL field has become more accessible in the latest years due to improvements in the understanding of machine-learning theory, coupled with the increased processing power from the better consumer-grade computer hardware. Within the full range of deep network architectures, Convolutional Neural Networks (CNN) gained the focus in recent DL studies. CNNs are multilayer neural networks capable of identifying patterns in data, both spatial and spectrally wise, and using these learned patterns for inference to classify data. Studies have shown they possess the ability to solve a wide variety of problems, such as text classification [1], image recognition [2], video analysis [3], and speech recognition [4].

Recent DL algorithms have found several applications in geoscience and Remote Sensing (RS) domains [5,6]. Studies have shown the feasibility and performance of CNNs for common RS tasks such as land cover classification [7–9], object detection [10,11], image pansharpening [12,13], change detection [14–17], and many others. These tasks have been carried successfully with a variety of RS image types, such as hyperspectral [7], multispectral [11,18], and in simple Red-Green-Blue (RGB) images [19].

Within the scope of land cover classification and change detection, the mapping of areas burnt by wildfires is crucial because of their ecological, social, and economic impacts [20]. The Brazilian Savannah (Cerrado) is one of the world's hotspots for biodiversity conservation and the world's richest neotropical savanna [21]. It is characterized by the regular occurrence of fires, either through natural or anthropogenic means, and approximately 170,000 km<sup>2</sup> of land within the region has burnt every year on average for the past ten years [22], making the Cerrado one of the regions most affected by wildfires globally [23].

Since the advent of satellite imagery, researchers have extensively attempted to map burnt areas as a critical step in understanding and preventing the social and environmental damage caused by fire [24]. Estimates show that fire consumes up to approximately 4 million km<sup>2</sup> of land yearly [25]. In the Cerrado biome, several studies analyze fire events based on remote sensing data, defining spatial patterns [26–28], temporal frequency [29,30], drivers of fire occurrences [31,32], and climatic effects [33].

Studies have found that commonly used global burnt area products such as the MCD64A1 MODIS dataset [34] often underestimate the real extent of the burnt areas [35–37]. Studies in the Brazilian territory [29,38] detected high values of errors in the MODIS-MCD45 product (commission error of 36.69% and omission error of 77.04%) and MODIS-MCD64 product (commission error of 45, 85% and omission error of 64.05%). Machine Learning (ML) algorithms have been shown to offer better results than such products [39]. Therefore, the current accuracy results of global fire mapping products show the need for advances in the detection and classification of the area of fires.

Shallow ML algorithms such as Support Vector Machines (SVM) and Random Forest (RF) along with shallow fully connected neural networks such as the Multilayer Perceptron (MLP) have been used to classify and detect burnt areas [39–44]. However, recently, these shallow algorithms were surpassed in most tasks by the deeper, more complex CNNs, of which the potential in classifying burnt area is still relatively unexplored. Studies that have applied DL algorithms to map burnt land have shown promising results [45,46], although several factors need further investigation, such as the use of different types of architectures and hyperparameter tuning. Several types of CNNs have been proposed and used for change detection and land cover classification [9,11,15,19,47]. The architectures known as Autoencoders have shown consistently good results among the many types of CNNs used for image segmentation [48–51]. Autoencoders use the concept of downsampling and upsampling feature maps, which makes them very efficient memory-wise and helps detect both high-level semantic information and low-level spatial detail. Therefore, Autoencoders offer a good choice of architecture to classify RS data given how memory intensive it can be and how it is highly dependent on spatial information.

Within the scope of segmentation in Remote Sensing images, the sampling technique has been a topic of discussion. The sampling process is generally performed by either placing random sampling windows or by sliding a sampling window along the image to collect pixel data in the form of smaller patches [52]. The sliding window technique is much more common, although there is no consensus on the optimal window size, which seems to depend on the type of image used and the target analyzed. Varying window sizes have been investigated in literature for different objects, such as (a) 17 × 17 pixels to detect oil palm trees in a plantation area from QuickBird image [53]; (b) 50 × 50 pixels to detect vehicles in aerial images [54], (c) 224 × 224 pixels for the analysis of damaged buildings using aerial images of 0.5-m resolution [55]; (d) 256 × 256 pixels to classify urban buildings from an image with 0.075-m resolution [56]; and (e) 400 × 400 pixels resampled to 256 × 256 pixels to classify urban land cover using high-resolution aerial images [57]. Although studies consider that the adequate window size should cover the intended target, a window size sensitivity analysis allows for the detection of the

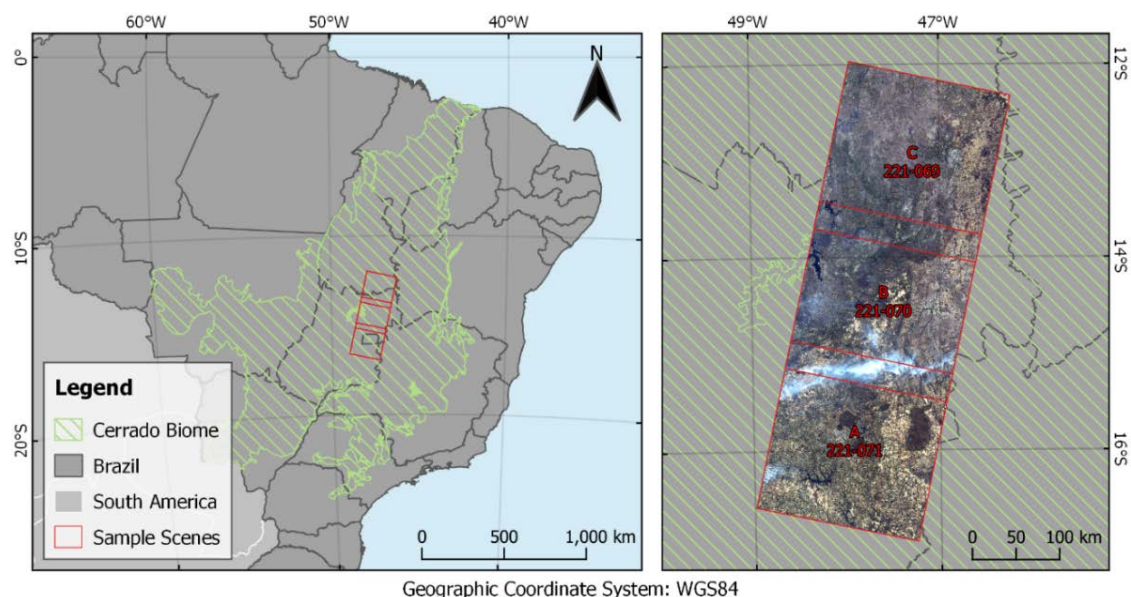
optimal dimension. A study mapping land cover using RapidEye images [58] tested different window sizes (5, 10, 15, 20, 25, 30, 35, and 40) and determined the dimension of  $30 \times 30$  pixels as the ideal size. Another study evaluated different window sizes (60, 80, 100, 120, 140, 160, 180, and 200) to locate cars in unmanned aerial vehicle (UAV) images [59], concluding that a patch size of  $160 \times 160$  pixels provided the best total accuracy. Additionally, studies show that some degree of overlap between windows is beneficial to the classification as it reduces the loss of contextual information along image patch borders [60].

This study aimed to investigate the use of DL algorithms to map burnt area changes within the Cerrado region to provide an accurate automated classification method. This research evaluated three CNN models based on the concept of the Autoencoder architecture: (a) the basic Autoencoder, (b) the U-Net, and (c) the ResUnet architectures, which propose improvements over the basic Autoencoder. Furthermore, we tested four sampling strategies to find optimal sampling window sizes for this specific classification task. In the following sections, we describe the study area; our dataset structure, how the models were built, how they were evaluated and lastly, we present the results found and a brief discussion over them.

## 2. Methodology

### 2.1. Landsat Data

Our training and testing datasets were created by collecting Tier 1 atmospherically corrected reflectance data detected by the Landsat 8 OLI+ sensors and pre-processed by the United States Geological Survey (USGS) agency. This study used bands 2 to 7, which offer the majority of the spectral information relevant to the detection of burnt lands for the same 30-meter spatial resolution. The training used the Landsat scenes (path-row) 221-71 and 221-70 (sites A and B), while the validation used scene 221-69 (site C) (Figure 1). The areas fit into the Cerrado biome and offer detection dates on the same day. The overlapping region between scenes B and C was excluded from scene B to avoid sharing data between training and validation.



**Figure 1.** Delimitation of the (A,B) training and (C) validation sites within the Brazilian Cerrado region.

To detect changes, we selected four different dates in August and September 2017 (August 9 and 25, September 10 and 26). The choice of date was based on the more significant occurrence of fires in the region during the end of the dry season [37]. Additionally, these dates offered the least amount of cloud cover throughout the year.

## 2.2. Burnt Area Change Mask

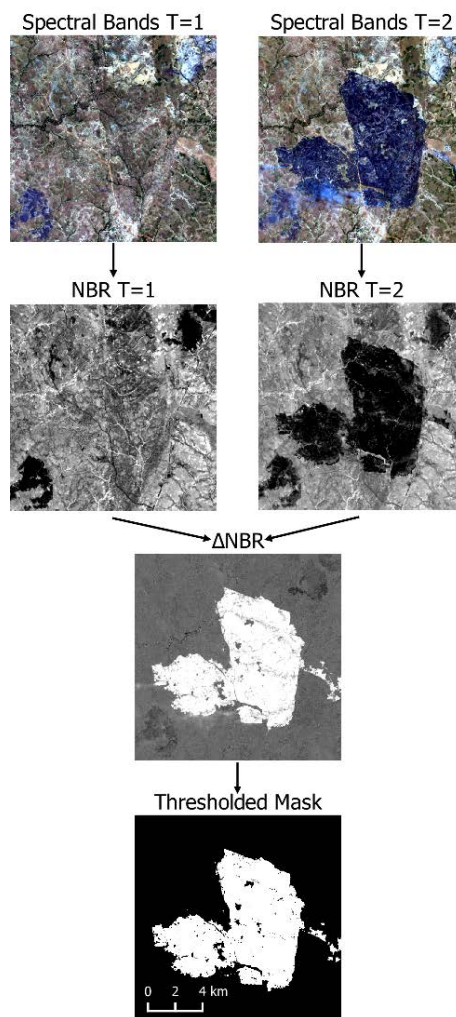
The elaboration of the ground truth mask used the Normalized Burn Ratio (*NBR*) spectral index (Equation (1)), which has been extensively used in research to highlight burnt areas and assess burn severity [38,61,62].

$$NBR = \frac{NIR - SWIR}{NIR + SWIR}, \quad (1)$$

where *NIR* and *SWIR* are the near and shortwave infrared bands, respectively. The *NBR* temporal difference ( $\Delta NBR$ ) can then be calculated to further highlight the burnt areas (Equation (2)).

$$\Delta NBR = NBR_{T1} - NBR_{T2}, \quad (2)$$

where *T1* and *T2* are the pre-fire and post-fire images, respectively. Specific  $\Delta NBR$  threshold values allow assessing the severity of the burn. In this study, we classified pixels with  $\Delta NBR$  values above 0.1 as burnt areas, regardless of the severity (Figure 2). Common false positives such as bodies of water and shadows, which are often also highlighted by this approach, were manually removed from the masks to guarantee that only burnt areas were present. This approach only detected new-burnt areas between two consecutive images without accounting for the accumulated burnt area.

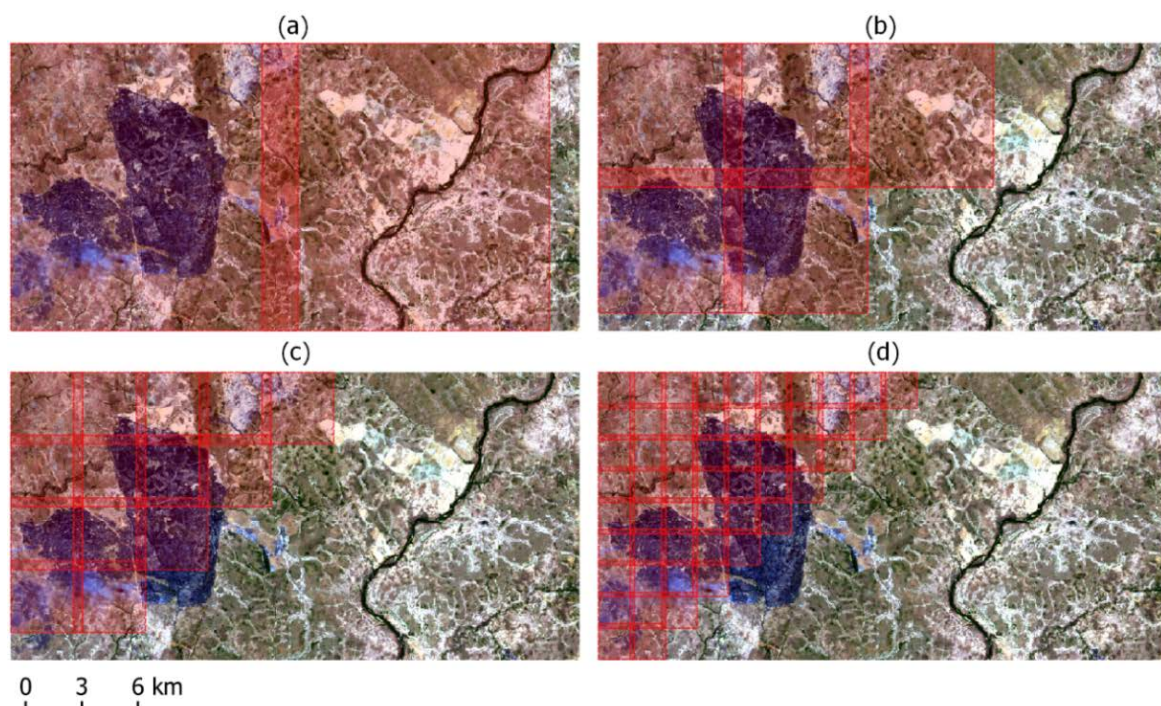


**Figure 2.** Visualization of the Normalized Burn Ratio thresholding process to generate the burnt area masks. From top to bottom, the true color satellite images, the calculated *NBR* values, the difference *NBR* between *T1* and *T2*, and the thresholded burnt area mask.

### 2.3. Data Structure

In this study, we used a bi-temporal approach in order to detect burnt area change. Therefore, images on two consecutive dates were paired and stacked depth-wise, generating a 12-band file associated with the respective change mask. Given our available images, this process generated three sets of bi-temporal images for each of the Landsat scenes.

Additionally, since data are structured in a batch by batch basis for deep learning models, our images had to be restructured and sampled as a 4D tensor containing multiple image patches and with shape  $[S \times H \times W \times B]$  where  $S$  is the number of samples,  $H$  and  $W$  the height and width of the patches in number pixels and  $B$  the number of bands in the bi-temporal image pair. In this study, we sampled the images through a sliding window of four different sizes based on power of two ( $2^n$ ) numbers: (a) 512 by 512, (b) 256 by 256, (c) 128 by 128, and (d) 64 by 64 pixels (Figure 3).



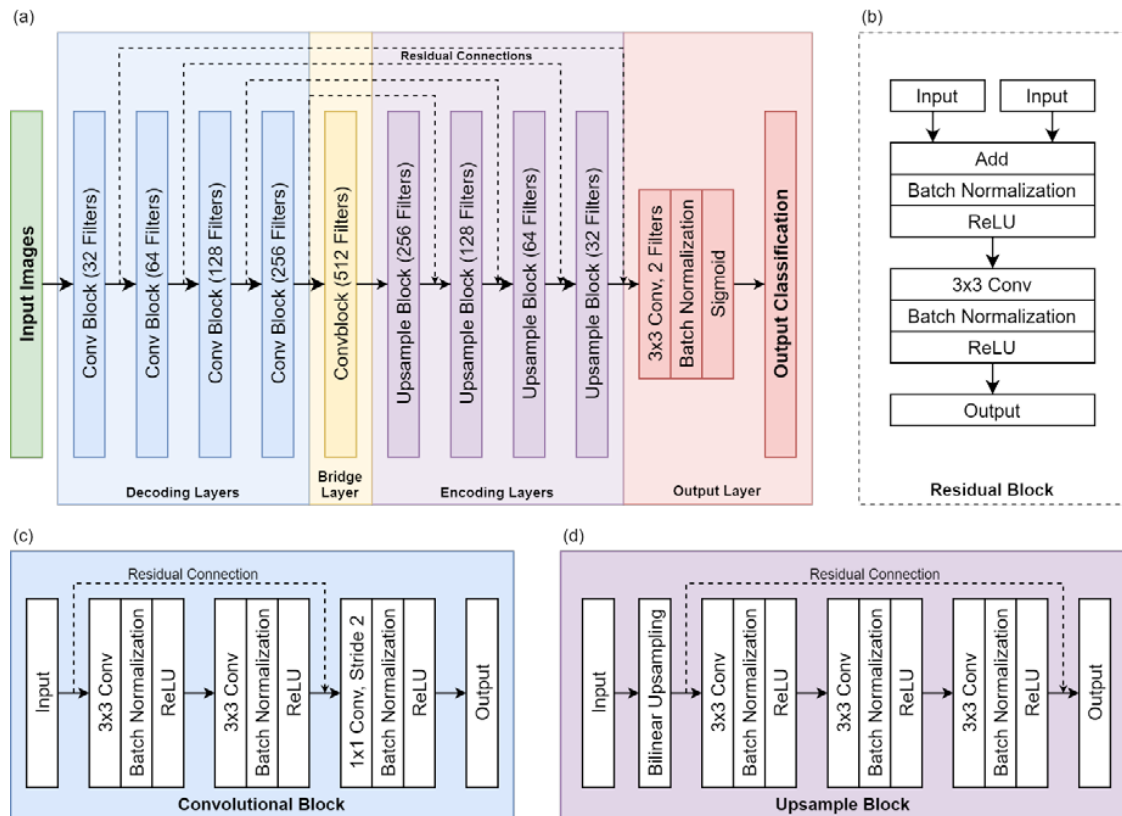
**Figure 3.** Example of the four different overlapping window sizes used to sample our images for the deep learning framework. Height and width in pixels of (a) 512, (b) 256, (c) 128 and (d) 64, respectively.

Furthermore, we used a 12.5% overlap between sampling windows to reduce the loss of predictive power near sample edges, an effect that is induced by the padding operation in convolutional layers and by the lack of contextual information near the patch edges. Incomplete windows, i.e., with empty pixel values, were discarded. The total number of samples per image generated through this process was 168, 747, 3140, and 12,860, respectively.

### 2.4. Deep Learning Models

The basis for the models used in this study was the Autoencoder model, which consists of an architecture that downsamples (decodes) the feature maps generated through convolutional layers to learn features compactly and then upsamples (encodes) them back to the desired output size. This process usually leads to the loss of spatial information as the feature maps are downsampled. The U-Net architecture [63] can be considered an evolution of the basic Autoencoder model, which tries to correct the loss of spatial information through the introduction of residual connections that propagate the information before being downsampled towards the upsampling layers. This allows the model to learn low-level detail while also keeping the high-level semantic information. A further

enhancement of the architecture has been proposed through the insertion of residual connections within the architecture's blocks, resulting in what has been called ResUnet [51,64]. These three architectures have been used to classify remote sensing data before with good results [48–51,64]. We adapted and evaluated these three architectures to describe possible differences when used to detect burnt area changes. Figure 4 describes the general structure of the models used.



**Figure 4.** Illustration of the structures for the architectures used in this study. (a) The full general architecture and the (b) residual block, (c) convolutional block, and (d) the upsampling block structures. The use of residual connections determines the model, where the Autoencoder, U-Net, and ResUnet use: no connections, connections between blocks, and connections between and within blocks, respectively.

In this study, the architecture of the three models have the same number of layers and basic structure. Still, they differ in the way the residual connections are used: (a) the Autoencoder uses no connections at all, (b) the U-Net architecture uses connections only between blocks from both sides of the structure, and (c) the ResUnet uses connections between and within the blocks. This allowed us to evaluate the effect of the addition of the residual connections. The Keras [65] Python framework was used to build and train the models and to classify the images.

## 2.5. Model Training

The three models shared most of the training parameters. To compute loss, we used the sum of the Binary Cross Entropy (BCE) loss and the Dice loss [66] functions (Equations (3)–(5)).

$$\text{Final Loss} = \text{BCE loss} + \text{Dice loss}, \quad (3)$$

where:

$$\text{BCE loss} = \frac{1}{m} \sum_{i=1}^m -(y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)), \quad (4)$$

$$Dice\ loss = 1 - \frac{1}{m} \sum_{i=1}^m \frac{2 \times \sum (y_i \times p_i)}{\sum y_i + \sum p_i} \quad (5)$$

where  $m$  is the number of mini-batches,  $y$  are the ground truth class values,  $y$  are the class scores from the sigmoid activation, and  $p$  are the predicted class values. The Dice loss comes from the Dice coefficient, also known as  $F1$ , which is especially useful for classifications with uneven class distributions (as in this study). This coefficient equally values positive and negative cases without the need to set arbitrary weights.

The gradient descent optimization used the RMSprop algorithm with a learning rate of  $10^{-4}$  that automatically decreased by a magnitude of 10 every time the loss reached a plateau, to a minimum of  $10^{-6}$ . The models were trained with the data from scenes A and B for a total of 200 epochs, which was enough to stabilize model loss and error for every model instance. The main differing parameter for model training was the batch size, which varied depending on the sample size as the available hardware memory constrained it. The batch sizes used were of 4, 8, 16, and 32, respectively, for the window sizes of 512, 256, 128, and 64. Those were the largest possible batch sizes that allowed us to fit the samples into memory for their respective window sizes without reducing the number of samples. The model training used a computer equipped with Nvidia GeForce RTX 2080 TI graphics card with 11 GB of GPU memory, 16 GB of RAM memory, and an Intel Core i7-4770K CPU.

## 2.6. Model Evaluation

As mentioned before, model validation used the data from scene C. The three main metrics used to evaluate the models were the  $F1$  measure, the  $Kappa$  coefficient, and the mean Intersection Over Union ( $mIoU$ ) value, represented by Equations (6), (9), and (12), respectively.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

where:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (7)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (8)$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (9)$$

where  $p_o$  is the rate of agreement between the ground truth and the classification, and  $p_e$  is the expected rate of random agreement (Equations (10) and (11)):

$$p_o = \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives} \quad (10)$$

$$p_e = \frac{(TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN)}{(TP + FN + TF + FP)^2} \quad (11)$$

$$mIoU = \frac{IoU_1 + IoU_2 \dots + IoU_n}{n} \quad (12)$$

where  $IoU$  is the area of the intersection divided by the area of the union of the classification and ground truth for a class, and  $n$  is the total number of classes. All three of these measures range from 0 to 1, where a result of 1 would represent a perfect classification. In this study, they provide a better quantitative assessment over the traditional accuracy value, which tends to be misleadingly optimistic in classifications with an imbalanced number of observations and a large number of background (negative) cases relative to the foreground (positive) cases [67,68].

In addition, we employed McNemar’s test [69] to evaluate whether the models were significantly different between each other. This test is a non-parametric test that mainly evaluates whether the error distribution between two classifications is similar. In this study, we used the variation of the test based on a chi-square distribution with a single degree of freedom and continuity correction [70] (Equation (13)).

$$\chi^2 = \frac{(|f_{12} - f_{21}| - 1)^2}{f_{12} + f_{21}} \quad (13)$$

where  $f_{12}$  and  $f_{21}$  are the frequency of observations in disagreement between two classifications in a contingency table. A  $p$ -value of 0.05 was used as the threshold value, where lower values indicate that the distributions between two compared models are significantly different.

### 3. Results

Table 1 lists the detailed validation results for the  $F1$ ,  $Kappa$  and  $mIoU$  measures for each model, while Figure 5 shows a visual comparison of these measures. The basic Autoencoder architecture showed the worst results overall, although it showed good  $F1$ ,  $Kappa$ , and  $mIoU$  values.

The U-Net and ResUnet architectures showed similar results. Despite performing better on average, the U-Net showed worse results in the time sequences of 08-25 to 09-10 and 09-10 to 09-26.

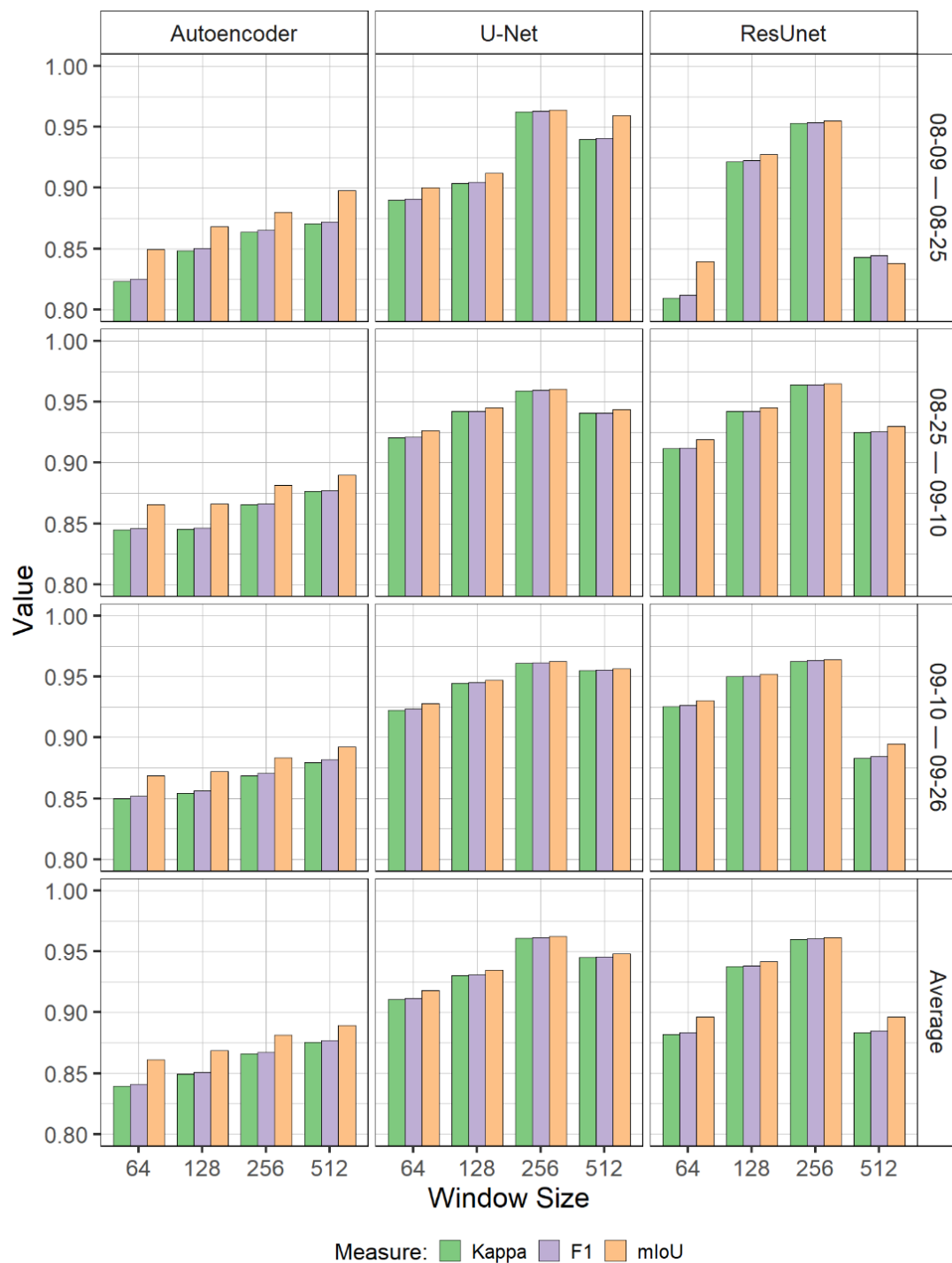
The models trained with samples with a size of 64 by 64 showed the worst results overall. The basic Autoencoder was the only model to show an improvement still when the window size was increased to 512 by 512. The ResUnet model showed a more marked loss of performance, increasing the window size to 512 by 512. Using a window size of 256 by 256 resulted in the best  $F1$ ,  $Kappa$ , and  $mIoU$  values for both the U-Net and ResUnet models.

In most cases, the models produced more false positives than false negatives (Figure 6). Improvements in the performance measures seemed to stem mainly from decreases in the number of false-positive predictions as the window sizes grew to 256 by 256. Comparatively, the number of false negatives varied little. However, the ResUnet model showed a noticeable increase in false negatives with the window size of 512 by 512. The time sequence between 08-25 and 09-10 showed the lowest amount of incorrectly classified pixels, which is explained by the fact that this sequence also showed the lowest extent of burnt areas overall.

**Table 1.** Evaluation metrics for each instance of the models separated by the time sequence and window sizes. Best results in each column highlighted in bold text.

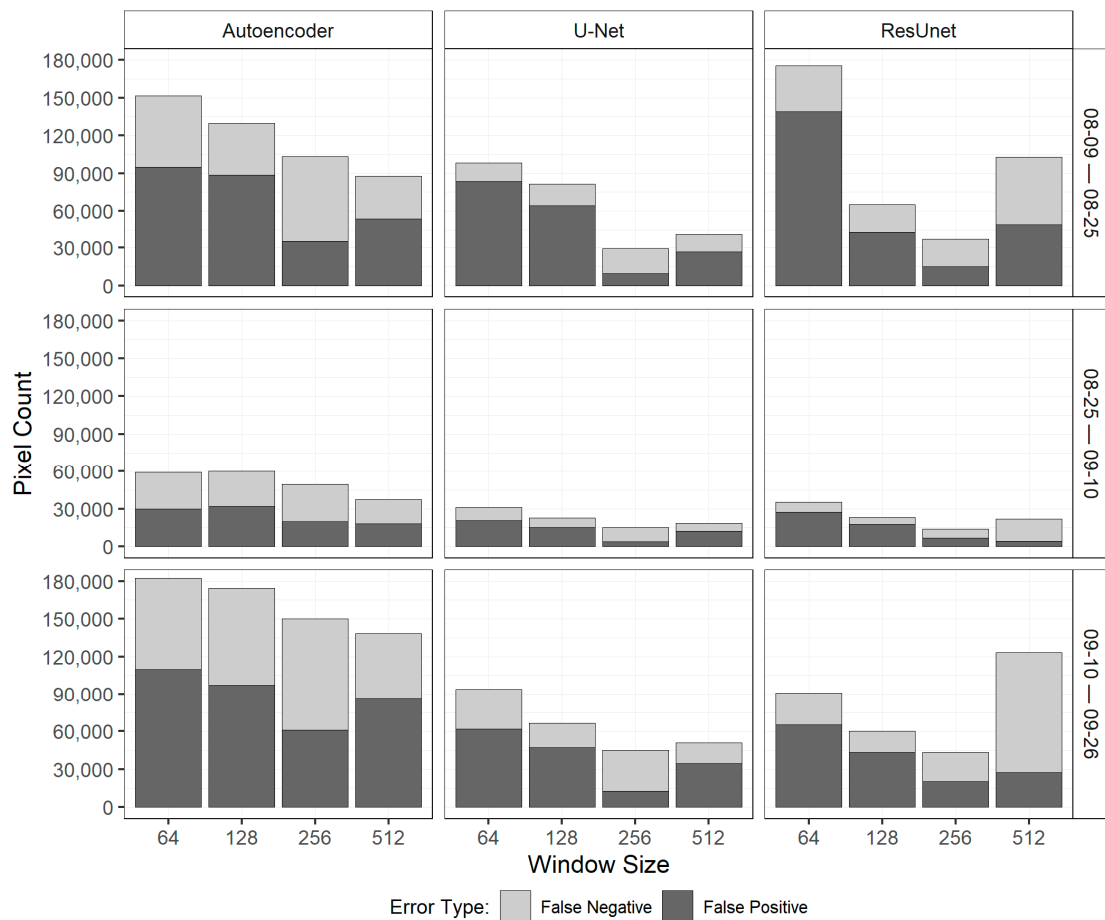
Model	08-09 to 08-25			08-25 to 09-10			09-10 to 09-26			Average		
	$Kappa$	$F1$	$mIoU$	$Kappa$	$F1$	$mIoU$	$Kappa$	$F1$	$mIoU$	$Kappa$	$F1$	$mIoU$
Autoencoder <sub>64</sub>	0.823	0.825	0.849	0.844	0.845	0.865	0.849	0.851	0.868	0.839	0.840	0.861
Autoencoder <sub>128</sub>	0.848	0.850	0.868	0.845	0.846	0.865	0.854	0.856	0.872	0.849	0.850	0.868
Autoencoder <sub>256</sub>	0.863	0.865	0.879	0.865	0.866	0.881	0.868	0.870	0.883	0.865	0.867	0.881
Autoencoder <sub>512</sub>	0.870	0.872	0.885	0.876	0.877	0.889	0.879	0.881	0.892	0.875	0.877	0.889
U-Net <sub>64</sub>	0.889	0.890	0.900	0.920	0.920	0.926	0.922	0.923	0.927	0.910	0.911	0.918
U-Net <sub>128</sub>	0.903	0.904	0.912	0.942	0.942	0.945	0.944	0.945	0.947	0.930	0.930	0.934
U-Net <sub>256</sub>	<b>0.962</b>	<b>0.963</b>	<b>0.964</b>	0.959	0.959	0.960	0.960	0.961	0.962	<b>0.960</b>	<b>0.961</b>	<b>0.962</b>
U-Net <sub>512</sub>	0.939	0.940	0.943	0.940	0.940	0.943	0.954	0.955	0.956	0.945	0.945	0.948
ResUnet <sub>64</sub>	0.809	0.811	0.839	0.911	0.912	0.918	0.925	0.926	0.930	0.882	0.883	0.896
ResUnet <sub>128</sub>	0.921	0.922	0.927	0.942	0.942	0.945	0.950	0.950	0.952	0.937	0.938	0.941
ResUnet <sub>256</sub>	0.953	0.953	0.955	<b>0.963</b>	<b>0.964</b>	<b>0.965</b>	<b>0.962</b>	<b>0.963</b>	<b>0.964</b>	0.959	0.960	0.961
ResUnet <sub>512</sub>	0.843	0.844	0.864	0.924	0.925	0.930	0.882	0.884	0.894	0.883	0.884	0.896





**Figure 5.** Visual comparison of the performance measures between the trained models for each time sequence and the average values with varying sample window sizes.

McNemar’s test shows that when compared, most models have significantly different error distributions, which means the observed differences in the results did not occur at random (Table 2). The only models found to be statistically similar were the 256-window U-Net and ResUnet. The differences in error distributions are also visually noticeable in the classification maps, particularly at the edges of burnt area patches. While the basic Autoencoder misclassified large groups of pixels, the U-Net and ResUnet models mostly showed misclassifications as very small groups or single pixels.

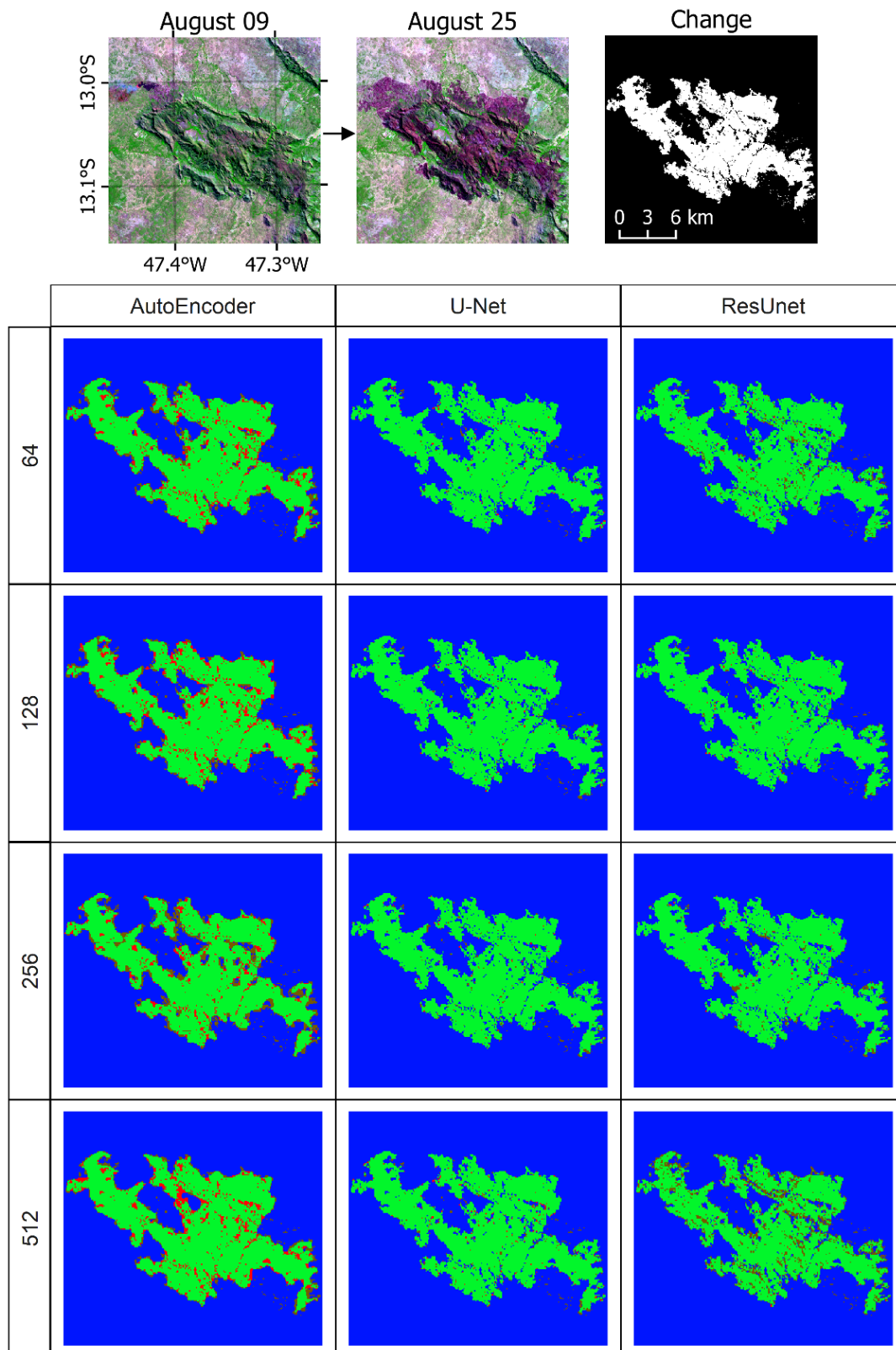


**Figure 6.** Error distributions in number of pixels for all model instances by time sequence and sample window size.

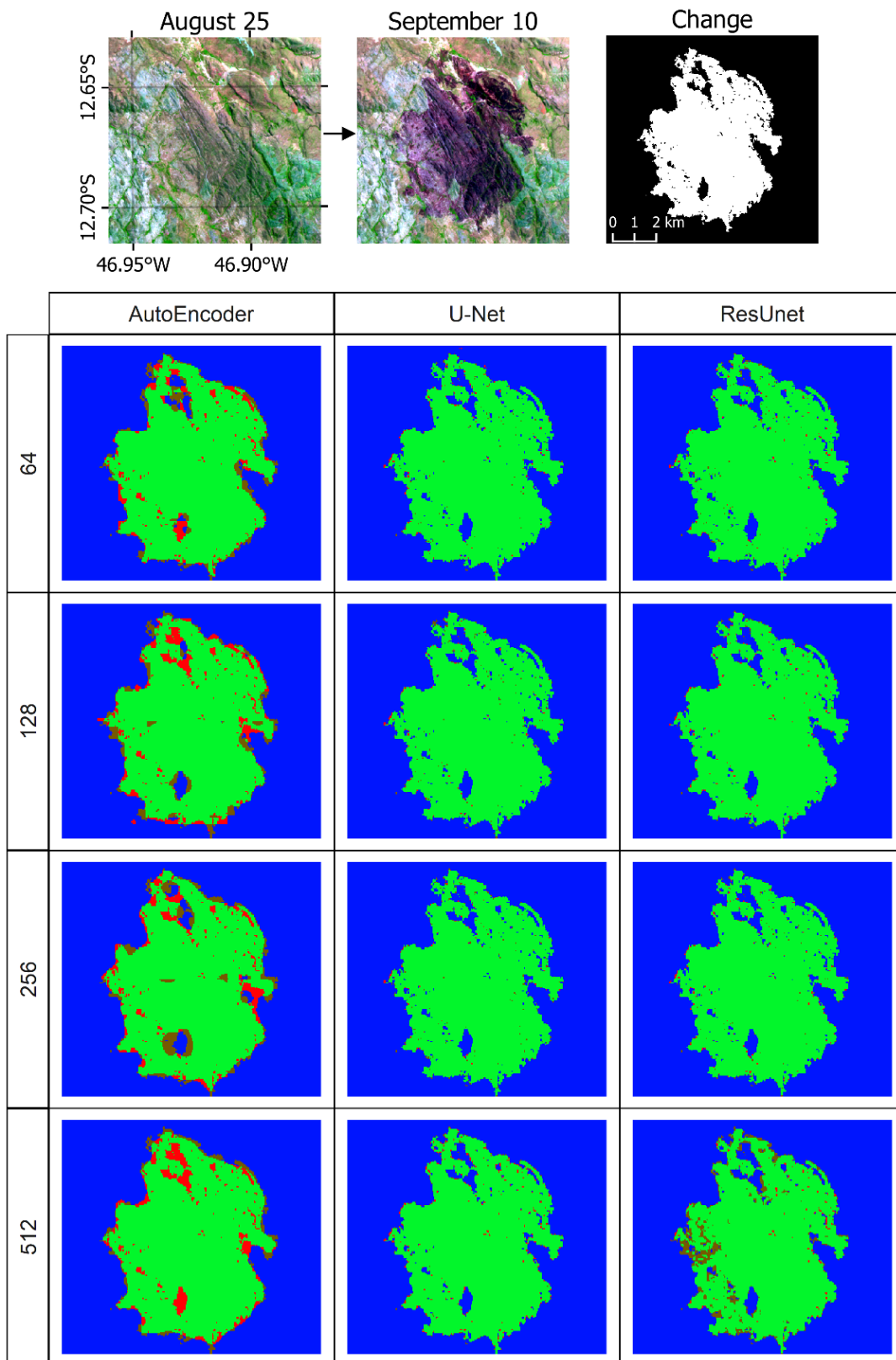
**Table 2.** *p* values of the McNemar’s test for comparing model classifications. Values under 0.05 indicate the error distribution from the two compared models are significantly different. (bold text indicates models that were statistically similar).

Model/Window	Autoencoder				U-Net				ResUnet			
	64	128	256	512	64	128	256	512	64	128	256	512
Autoencoder	64											
	128	<0.001										
	256	<0.001	<0.001									
	512	<0.001	<0.001	<0.001								
U-Net	64	<0.001	<0.001	<0.001	<0.001							
	128	<0.001	<0.001	<0.001	<0.001	<0.001						
	256	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001					
	512	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001				
ResUnet	64	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001			
	128	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001		
	256	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<b>0.089</b>	<0.001	<0.001	<0.001	
	512	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

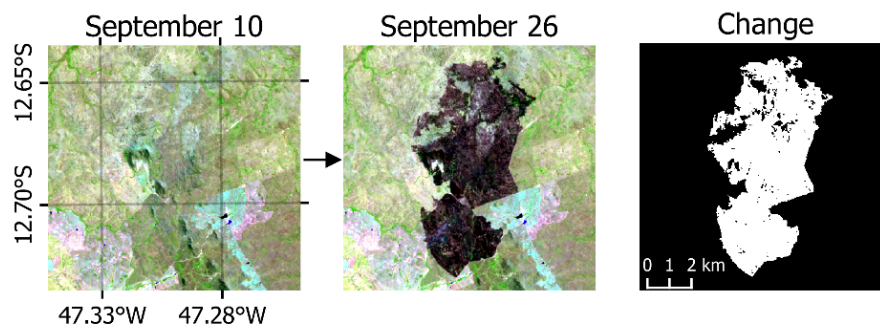
There were no noticeable differences between time sequences despite the physical and phenological changes (Figures 7–9). The models automatically masked water bodies and most shadows, which are spectrally similar to burnt areas. However, burnt area patches with cloud cover in either image in a sequence were still sources of error in the classifications. Despite that, cloud shadows in unburnt lands were still correctly classified as negatives in most cases. Both the U-Net and ResUnet models were able to classify unclouded patches of burnt land with a low occurrence of errors.



**Figure 7.** Example of a classified burnt area patch of the 9 August to 25 August sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.



**Figure 8.** Example of a classified burnt area patch of the 25 August to 10 September sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.



	AutoEncoder	U-Net	ResUnet
64			
128			
256			
512			

Legend: ■ True Positive ■ False Positive ■ True Negative ■ False Negative

**Figure 9.** Example of a classified burnt area patch of the 10 September to 26 September sequence. On top, the false-color Landsat images (R: band 6, G: band 5 and B: band 4) along with the change mask and on the bottom the model classifications coded by prediction type.

#### 4. Discussion

Results have shown that the DL models evaluated offer excellent classification results when used to detect burnt area changes. Even the worst model using the Autoencoder architecture with a sample window of 64 by 64 resulted in  $F1$ ,  $Kappa$ , and  $mIoU$  values over 0.8, which can already be considered a good result. The addition of residual connections between the decoding and encoding layers in the U-Net significantly improved the results. In contrast, the ResUnet's addition of connections within individual blocks gave marginal improvements and only in some cases. Visually, the Autoencoder's lack of connections translated to a noticeable loss of spatial information in the form of less detailed contours between the positive and negative classes. Overall, the U-Net architecture showed the best results, although not much higher than the ResUnet architecture, which was superior in the time sequences with greater extensions of burnt land. Regions with clouded patches of burnt land were among the primary sources of errors in the classifications as occasionally, the models misclassified the cloud shadows as extensions of the burnt land patches, generating false positives. Despite that, the models correctly classified regions without mixed burnt areas and cloud shadows and automatically masked objects commonly detected as false positives through  $\Delta NBR$  thresholding, therefore reducing the need for human intervention. The presence of cloud cover is one of the main limitations when using Landsat data for change detection as it is a common occurrence that impacts both the creation of a ground truth mask and the training of the models. Radar data can be used instead but at the cost of a significant loss of spectral information and possibly accuracy [71]. Studies have been carried using CNNs and Synthetic Aperture Radar (SAR) data to detect burnt areas with results similar to those found in this study [46,72] although to much smaller extents. Our bi-temporal approach was similar to that used by the Brazilian Institute of Space Research (INPE) to produce official burnt area reports [73]. However, our use of DL architectures instead of a thresholding process produced a much lower rate of false positives (commission errors) and false negatives (omission errors). Furthermore, DL models can be trained incrementally with new training data and further improve results, although up to a specific limit.

Increasing the sample window size improved the results despite simultaneously decreasing the training batch sizes, although only up to 256 by 256 in the case of the U-Net and ResUnet models. The size of 512 by 512 worsened the results, particularly for the ResUnet model, which showed results close to the same model using the 64 by 64 samples. The cause can be attributed to the lower batch size used. Studies have shown that smaller batch sizes can introduce more noise in the training gradients, leading to a loss of generalizability and, therefore, less accuracy [74,75]. However, the window size of 64 by 64 showed the worst results overall, even using the largest batch size, which shows that there is possibly a balance between sample window size and batch size. This problem is ultimately limited by the quantity of memory available in the graphics card, which determines the number of samples, the size of samples, and the batch size that can be used in the same training process. In addition, increasing the model complexity (e.g., by increasing the number of layers or filters) can exponentially increase the memory required for training. Remote sensing data can be highly memory intensive, especially at higher spatial and spectral resolutions, making the process of optimizing the training parameters for DL models challenging with consumer-grade hardware.

The loss of performance with smaller window sizes is also related to the size of the object at hand. While small window sizes might cover the full extent of small objects, they cannot fully cover larger objects, leading to less information about the relationship between the object of study and its surroundings. Given the way convolutional networks function, the information within each image patch is highly important. In this study, the extent of burnt areas ranged from single pixels (900 m<sup>2</sup>) to several square kilometers, and, as seen in Figure 3, the smaller window sizes created several image patches with low or no background-foreground context, i.e., without enough information about the burnt area border dynamics. Despite that, the results show that the sampling window does not necessarily need to cover the full extent of the object of detection, corroborating with results found in other studies [59].

## 5. Conclusions

In this study, we evaluated three Deep Learning models to detect burnt area changes in three bi-temporal Landsat image pairs: a basic Autoencoder, U-Net, and ResUnet. All three networks were based on the same principles but with differences in the use of residual connections. The training and validation of the models used Landsat data from scenes within the region of the Brazilian Cerrado. The models were trained with four different sample window sizes in pixels to verify performance differences: 64 by 64, 128 by 128, 256 by 256, and 512 by 512.

Results have shown that the architectures used are a reliable automated way to map burnt area changes between bi-temporal image pairs in the Cerrado. However, the U-Net and ResUnet models were superior to the basic Autoencoder as the introduction of residual connections significantly improved the results. The sample window size of 256 by 256 pixels showed the best results for the U-Net and ResUnet models, and further increasing it produced worse results for both of these models. The model evaluation considered the *F1*, *Kappa*, and *mIoU* measures, of which the 256 by 256 window U-Net model achieved the best overall results with average values of 0.960, 0.961, and 0.962, respectively. The ResUnet model had slightly worse results on average, but slightly better results in two of the three time sequences evaluated. McNemar's test verified the possibility that the differences between classifications were not statistically significant, and only the U-Net and ResUnet models using  $256 \times 256$ -pixel samples were found to be similar, while every other model was statistically unique.

The Cerrado biome is an important region given its biodiversity, but it is constantly under the threat of destruction through fires and deforestation. We recommend that future studies investigate more uses of current Deep Learning techniques to provide better solutions for the detection and mitigation of these threats. In addition, certain spectral vegetation and burn indexes that have been shown to possibly improve the detection of burnt land [76,77] were not used in this study and could be investigated in future works. A few other suggestions for further studies of the theme arose from certain limitations found in this study. We recommend an investigation of the effect of training batch sizes along with sample window sizes. The batch size is an essential factor in the model performances but is limited by memory, along with several other parameters relevant to DL models. Secondly, the presence of cloud cover in Landsat images is a source of error. Radar data partly solves this problem at the expense of spectral information. Therefore, the possibility of the use of mixed sensor data for burnt area mapping should be investigated. Another future investigation should be to compare the performance of DL algorithms with shallow ML algorithms to highlight differences in performance.

**Author Contributions:** Conceptualization, P.P.d.B. and O.A.d.C.J.; methodology, P.P.d.B. and O.A.d.C.J.; software P.P.d.B., O.A.d.C.J. and O.L.F.d.C.; validation, P.P.d.B.; writing—original draft preparation, P.P.d.B. and O.A.d.C.J.; writing—review and editing, P.P.d.B. and O.A.d.C.J.; supervision, O.A.d.C.J.; project administration, O.A.d.C.J., R.A.T.G., R.F.G.; funding acquisition, O.A.d.C.J., R.A.T.G., R.F.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** The following institutions funded this research: National Council for Scientific and Technological Development (434838/2018-7), Coordination for the Improvement of Higher Education Personnel and the Union Heritage Secretariat of the Ministry of Economy.

**Acknowledgments:** The authors are grateful for financial support from CNPq fellowship (Osmar Abílio de Carvalho Júnior, Roberto Arnaldo Trancoso Gomes, and Renato Fontes Guimarães). Special thanks are given to the research group of the Laboratory of Spatial Information System of the University of Brasilia for technical support. The authors thank the researchers from the Union Heritage Secretariat of the Ministry of Economy, who encouraged research with deep learning. This study was financed in part by the Coordination for the Improvement of Higher Education Personnel (CAPES)—Finance Code 001. Finally, the authors acknowledge the contribution of anonymous reviewers.

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

## References

- Zhang, X.; Zhao, J.; LeCun, Y. Character-level Convolutional Networks for Text Classification. In *Advances in Neural Information Processing Systems 28*; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; Curran Associates Inc.: Montreal, QC, Canada, 2015; pp. 649–657.
- Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
- Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [[CrossRef](#)]
- Abdel-Hamid, O.; Mohamed, A.; Jiang, H.; Penn, G. Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 4277–4280.
- Zhang, L.; Zhang, L.; Kumar, V. Deep learning for Remote Sensing Data. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
- 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](#)]
- Mou, L.; Ghamisi, P.; Zhu, X.X. Fully conv-deconv network for unsupervised spectral-spatial feature extraction of hyperspectral imagery via residual learning. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; Volume 56, pp. 5181–5184.
- Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
- Scott, G.J.; England, M.R.; Starns, W.A.; Marcum, R.A.; Davis, C.H. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [[CrossRef](#)]
- Imamoglu, N.; Kimura, M.; Miyamoto, H.; Fujita, A.; Nakamura, R. Solar Power Plant Detection on Multi-Spectral Satellite Imagery using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion. *arXiv* **2017**, arXiv:1704.06410.
- Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *Int. J. Comput. Intell. Appl.* **2017**, *16*, 1750001. [[CrossRef](#)]
- Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A Multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 978–989. [[CrossRef](#)]
- Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, 1–15. [[CrossRef](#)]
- Alcantarilla, P.F.; Stent, S.; Ros, G.; Arroyo, R.; Gherardi, R. Street-view change detection with deconvolutional networks. *Auton Robot* **2018**, *42*, 1301–1322. [[CrossRef](#)]
- Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 125–138. [[CrossRef](#)] [[PubMed](#)]
- Zhang, P.; Gong, M.; Su, L.; Liu, J.; Li, Z. Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 24–41. [[CrossRef](#)]
- Zhao, J.; Gong, M.; Liu, J.; Jiao, L. Deep learning to classify difference image for image change detection. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; IEEE: Beijing, China, 2014; pp. 411–417.
- Zeng, X.; Yang, J.; Deng, X.; An, W.; Li, J. Cloud detection of remote sensing images on Landsat-8 by deep learning. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, 8 August 2018; Jiang, X., Hwang, J.-N., Eds.; SPIE: Shanghai, China, 2018; p. 173.
- Zhan, Y.; Wang, J.; Shi, J.; Cheng, G.; Yao, L.; Sun, W. Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1785–1789. [[CrossRef](#)]
- Chuvieco, E.; Aguado, I.; Yebra, M.; Nieto, H.; Salas, J.; Martín, M.P.; Vilar, L.; Martínez, J.; Martín, S.; Ibarra, P.; et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecol. Model.* **2010**, *221*, 46–58. [[CrossRef](#)]



21. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; da Fonseca, G.A.B.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858. [[CrossRef](#)]
22. INPE—Instituto Nacional de Pesquisas Espaciais Monitoramento de Queimadas. Available online: <http://www.inpe.br/queimadas> (accessed on 6 November 2017).
23. Costafreda-Aumedes, S.; Comas, C.; Vega-Garcia, C. Human-caused fire occurrence modelling in perspective: A review. *Int. J. Wildland Fire* **2017**, *26*, 983. [[CrossRef](#)]
24. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. [[CrossRef](#)]
25. Chuvieco, E.; Lizundia-Loiola, J.; Pettinari, M.L.; Ramo, R.; Padilla, M.; Tansey, K.; Mouillot, F.; Laurent, P.; Storm, T.; Heil, A.; et al. Generation and analysis of a new global burned area product based on MODIS 250m reflectance bands and thermal anomalies. *Earth Syst. Sci. Data* **2018**, *10*, 2015–2031. [[CrossRef](#)]
26. Daldegan, G.A.; de Carvalho Júnior, O.A.; Guimarães, R.F.; Gomes, R.A.T.; de Ribeiro, F.F.; McManus, C. Spatial patterns of fire recurrence using remote sensing and GIS in the Brazilian savanna: Serra do Tombador Nature Reserve, Brazil. *Remote Sens.* **2014**, *6*, 9873–9894. [[CrossRef](#)]
27. Pereira, J.M.C. Remote sensing of burned areas in tropical savannas. *Int. J. Wildland Fire* **2003**, *12*, 259. [[CrossRef](#)]
28. Sousa, I.M.P.; de Carvalho, E.V.; Batista, A.C.; Machado, I.E.S.; Tavares, M.E.F.; Giongo, M. Identification of burned areas by special index in a cerrado region of the state of tocantins, Brazil. *Floresta* **2018**, *48*, 553. [[CrossRef](#)]
29. De Carvalho Júnior, O.A.; Guimarães, R.F.; Silva, C.; Gomes, R.A.T. Standardized time-series and interannual phenological deviation: New techniques for burned-area detection using long-term modis-nbr dataset. *Remote Sens.* **2015**, *7*, 6950–6985. [[CrossRef](#)]
30. Pereira Júnior, A.C.; Oliveira, S.L.J.; Pereira, J.M.C.; Turkman, M.A.A. Modelling fire frequency in a cerrado savanna protected area. *PLoS ONE* **2014**, *9*, e102380. [[CrossRef](#)] [[PubMed](#)]
31. Alvarado, S.T.; Fornazari, T.; Cóstola, A.; Morellato, L.P.C.; Silva, T.S.F. Drivers of fire occurrence in a mountainous Brazilian cerrado savanna: Tracking long-term fire regimes using remote sensing. *Ecol. Indic.* **2017**, *78*, 270–281. [[CrossRef](#)]
32. De Bem, P.P.; de Carvalho Júnior, O.A.; Matricardi, E.A.T.; Guimarães, R.F.; Gomes, R.A.T. Predicting wildfire vulnerability using logistic regression and artificial neural networks: A case study in Brazil's Federal District. *Int. J. Wildland Fire* **2019**, *28*, 35. [[CrossRef](#)]
33. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognit.* **2017**, *61*, 539–556. [[CrossRef](#)]
34. Giglio, L.C.J. *MCD64A1 MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015.
35. Hall, J.V.; Loboda, T.V.; Giglio, L.; McCarty, G.W. A MODIS-based burned area assessment for Russian croplands: Mapping requirements and challenges. *Remote Sens. Environ.* **2016**, *184*, 506–521. [[CrossRef](#)]
36. Hawbaker, T.J.; Vanderhoof, M.K.; Beal, Y.-J.; Takacs, J.D.; Schmidt, G.L.; Falgout, J.T.; Williams, B.; Fairaux, N.M.; Caldwell, M.K.; Picotte, J.J.; et al. Mapping burned areas using dense time-series of Landsat data. *Remote Sens. Environ.* **2017**, *198*, 504–522. [[CrossRef](#)]
37. Moreira de Araújo, F.; Ferreira, L.G.; Arantes, A.E. Distribution patterns of burned areas in the Brazilian biomes: An analysis based on satellite data for the 2002–2010 period. *Remote Sens.* **2012**, *4*, 1929–1946. [[CrossRef](#)]
38. Santana, N.; de Carvalho Júnior, O.; Gomes, R.; Guimarães, R. Burned-area detection in Amazonian environments using standardized time series per pixel in MODIS data. *Remote Sens.* **2018**, *10*, 1904. [[CrossRef](#)]
39. Pereira, A.; Pereira, J.; Libonati, R.; Oom, D.; Setzer, A.; Morelli, F.; Machado-Silva, F.; de Carvalho, L. Burned area mapping in the Brazilian savanna using a one-class support vector machine trained by active fires. *Remote Sens.* **2017**, *9*, 1161. [[CrossRef](#)]
40. Ramo, R.; Chuvieco, E. Developing a random forest algorithm for MODIS global burned area classification. *Remote Sens.* **2017**, *9*, 1193. [[CrossRef](#)]
41. Mithal, V.; Nayak, G.; Khandelwal, A.; Kumar, V.; Nemani, R.; Oza, N. Mapping burned areas in tropical forests using a novel machine learning framework. *Remote Sens.* **2018**, *10*, 69. [[CrossRef](#)]

42. Al-Rawi, K.R.; Casanova, J.L.; Calle, A. Burned area mapping system and fire detection system, based on neural networks and NOAA-AVHRR imagery. *Int. J. Remote Sens.* **2010**, *22*, 2015–2032. [[CrossRef](#)]
43. Meng, R.; Zhao, F. Remote sensing of fire effects. A review for recent advances in burned area and burn severity mapping. In *Remote Sensing of Hydrometeorological Hazards*; Petropoulos, G.P., Islam, T., Eds.; CRC Press: Boca Raton, FL, USA, 2017; pp. 261–276.
44. Shan, T.; Wang, C.; Chen, F.; Wu, Q.; Li, B.; Yu, B.; Shirazi, Z.; Lin, Z.; Wu, W. A Burned Area Mapping Algorithm for Chinese FengYun-3 MERSI Satellite Data. *Remote Sens.* **2017**, *9*, 736. [[CrossRef](#)]
45. Langford, Z.; Kumar, J.; Hoffman, F. Wildfire Mapping in Interior Alaska Using Deep Neural Networks on Imbalanced Datasets. In *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 17–20 November 2018*; IEEE: Singapore, 2018; pp. 770–778.
46. Zhang, P.; Nascetti, A.; Ban, Y.; Gong, M. An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 50–62. [[CrossRef](#)]
47. Ba, R.; Chen, C.; Yuan, J.; Song, W.; Lo, S. SmokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention. *Remote Sens.* **2019**, *11*, 1702. [[CrossRef](#)]
48. De Bem, P.; de Carvalho Junior, O.; Fontes Guimarães, R.; Trancoso Gomes, R. Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. *Remote Sens.* **2020**, *12*, 901. [[CrossRef](#)]
49. Li, L. Deep residual autoencoder with multiscaling for semantic segmentation of land-use images. *Remote Sens.* **2019**, *11*, 2142. [[CrossRef](#)]
50. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-temporal SAR data large-scale crop mapping based on u-net model. *Remote Sens.* **2019**, *11*, 68. [[CrossRef](#)]
51. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [[CrossRef](#)]
52. Bermudez, J.D.; Happ, P.N.; Oliveira, D.A.B.; Feitosa, R.Q. Sar to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *IV-1*, 5–11. [[CrossRef](#)]
53. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* **2016**, *9*, 22. [[CrossRef](#)]
54. Yohei, K.; Hiroyuki, M. ryosuke shibasaki A CNN-based method of vehicle detection from aerial images using hard example mining. *Remote Sens.* **2018**, *10*, 124. [[CrossRef](#)]
55. Ma, H.; Liu, Y.; Ren, Y.; Wang, D.; Yu, L.; Yu, J. Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images. *Remote Sens.* **2020**, *12*, 260. [[CrossRef](#)]
56. Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network. *Remote Sens.* **2019**, *11*, 1774. [[CrossRef](#)]
57. Liu, C.; Zeng, D.; Wu, H.; Wang, Y.; Jia, S.; Xin, L. Urban land cover classification of high-resolution aerial imagery using a relation-enhanced multiscale convolutional network. *Remote Sens.* **2020**, *12*, 311. [[CrossRef](#)]
58. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sens.* **2018**, *10*, 1119. [[CrossRef](#)]
59. Ammour, N.; Alhichri, H.; Bazi, Y.; Benjdira, B.; Alajlan, N.; Zuair, M. Deep learning approach for car detection in uav imagery. *Remote Sens.* **2017**, *9*, 312. [[CrossRef](#)]
60. De Albuquerque, A.O.; de Carvalho Júnior, O.A.C.; de Carvalho, O.L.F.; de Bem, P.P.; Ferreira, P.H.G.; de dos Moura, R.S.; Silva, C.R.; Gomes, R.A.T.; Guimarães, R.F. Deep semantic segmentation of center pivot irrigation systems from remotely sensed data. *Remote Sens.* **2020**, *12*, 2159. [[CrossRef](#)]
61. Escuin, S.; Navarro, R.; Fernández, P. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (normalized difference vegetation index) derived from LANDSAT TM/ETM images. *Int. J. Remote Sens.* **2008**, *29*, 1053–1073. [[CrossRef](#)]
62. 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](#)]
63. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597.

64. Cao, K.; Zhang, X. An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sens.* **2020**, *12*, 1128. [[CrossRef](#)]
65. Chollet, F. Others Keras. Available online: <https://keras.io> (accessed on 6 July 2020).
66. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *arXiv* **2016**, arXiv:1606.04797.
67. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [[CrossRef](#)]
68. Maratea, A.; Petrosino, A.; Manzo, M. Adjusted F-measure and kernel scaling for imbalanced data learning. *Inf. Sci.* **2014**, *257*, 331–341. [[CrossRef](#)]
69. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [[CrossRef](#)]
70. Foody, G.M. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogram. Eng. Remote Sens.* **2004**, *70*, 627–633. [[CrossRef](#)]
71. Tanase, M.A.; Belenguer-Plomer, M.A.; Roteta, E.; Bastarrika, A.; Wheeler, J.; Fernández-Carrillo, Á.; Tansey, K.; Wiedemann, W.; Navratil, P.; Lohberger, S.; et al. Burned area detection and mapping: Intercomparison of sentinel-1 and sentinel-2 based algorithms over tropical Africa. *Remote Sens.* **2020**, *12*, 334. [[CrossRef](#)]
72. Ban, Y.; Zhang, P.; Nascetti, A.; Bevington, A.R.; Wulder, M.A. Near real-time wildfire progression monitoring with sentinel-1 SAR time series and deep learning. *Sci. Rep.* **2020**, *10*, 1322. [[CrossRef](#)] [[PubMed](#)]
73. Melchiori, A.E.; Setzer, A.W.; Morelli, F.; Libonati, R.; de Cândido, P.A.; de Jesús, S.C. A Landsat-TM/OLI algorithm for burned areas in the Brazilian Cerrado: Preliminary results. In *Advances in Forest Fire Research*; Imprensa da Universidade de Coimbra: Coimbra, Portugal, 2014; Volume 4, pp. 1302–1311. ISBN 978-989-26-0884-6.
74. Kandel, I.; Castelli, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express* **2020**, S2405959519303455. [[CrossRef](#)]
75. Radiuk, P.M. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. *Inf. Technol. Manag. Sci.* **2017**, *20*. [[CrossRef](#)]
76. Axel, A. Burned area mapping of an escaped fire into tropical dry forest in western madagascar using multi-season landsat oli data. *Remote Sens.* **2018**, *10*, 371. [[CrossRef](#)]
77. Saulino, L.; Rita, A.; Migliozzi, A.; Maffei, C.; Allevato, E.; Garonna, A.P.; Saracino, A. Detecting burn severity across mediterranean forest types by coupling medium-spatial resolution satellite imagery and field data. *Remote Sens.* **2020**, *12*, 741. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).