SEMANTIC SEGMENTATION OF BURNED AREAS IN SATELLITE IMAGES USING A U-NET-BASED CONVOLUTIONAL NEURAL NETWORK

A. K. Brand [1,*], A. Manandhar [1]

[1] RSS – Remote Sensing Solutions GmbH, Dingolfinger Str. 9, 81637 München – brand@remote-sensing-solutions.com

KEY WORDS: Deep Learning, Burned Area Mapping, Sentinel-2, Semantic Segmentation, Convolutional Neural Network, U-Net.

ABSTRACT:

The use of remote sensing data for burned area mapping hast led to unprecedented advances within the field in recent years. Although threshold and traditional machine learning based methods have successfully been applied to the task, they implicate drawbacks including the involvement of complex rule sets and requirement of previous feature engineering. In contrast, deep learning offers an end-to-end solution for image analysis and semantic segmentation. In this study, a variation of U-Net is investigated for mapping burned areas in mono-temporal Sentinel-2 imagery. The experimental setup is divided into two phases. The first one includes a performance evaluation based on test data, while the second serves as a use case simulation and spatial evaluation of training data quality. The former is especially designed to compare the results between two local (trained only with data from the respective research areas) and a global (trained with the whole dataset) variant of the model with research areas being Indonesia and Central Africa. The networks are trained from scratch with a manually generated customized training dataset. The application of the two variants per region revealed only slight superiority of the local model (macro-F1: 92%) over the global model (macro-F1: 91%) in Indonesia with no difference in overall accuracy (OA) at 94%. In Central Africa, the results of the global and local model are the same in both metrics (OA: 84%, macro-F1: 82%). Overall, the outcome demonstrates the global model's ability to generalize despite high dissimilarities between the research areas.

1. INTRODUCTION

Fire is a natural and ecologically relevant process in many ecosystems (Kelly & Brotons, 2017). However, following an intensification of anthropogenous land use, it has become an increasingly unpredictable socio-ecological hazard that effects many countries worldwide (Pereira et al., 2017). In Indonesia, forest and bush fires are often associated with land use conversion i.e. agricultural expansion and land conflict (Sizer et al., 2014). Converted peatland areas are especially prone to fires spreading out of control (Vetrita & Cochrane, 2020). Additional to agricultural fires that cause wildfires in Central Africa, the seasonal character of water availability in the savannah leads to the accumulation of easily ignited fuels that have the potential to burn every year (Pereira, 2003). Not only do these natural disasters in both regions largely contribute to pyrogenic emissions (Knorr et al., 2016), they also threaten biodiversity and forest health (Lewis et al., 2015) as well as the economy, lives and property (Chuvieco et al., 2010). With the impact of fires expected to increase (Roos et al., 2016), it is particularly relevant to monitor these events to generate knowledge about fire extent, location and frequency (Chuvieco et al., 2019). Burned area maps present a valuable option for devising prevention and recovery policies. As remote sensing data provides global information at high repetition rates, its utilization has led to unprecedented advances within the field of burned area mapping in recent years (Pereira et al., 2017).

A large variety of products and techniques has been developed for this purpose. Most prevalent in literature is the use of Landsat data in regional studies and MODIS data for the development of global burned area products. While they present an important source of information for multiple user communities, it has been shown that especially in areas prone to smaller fires, the low spatial resolution can lead to an underestimation of total burned area (van der Werf et al., 2017). Consequently, there is an

outstanding demand for high-resolution burned area products that improve these estimates. Sentinel-2 data is well suited for this task. With a combined revisit time of 5 days considering both missions (Sentinel-2A and Sentinel-2B) and the included Near Infrared (NIR) and Shortwave Infrared (SWIR) spectral bands that are especially sensitive to fire effects (Pleniou & Koutsias, 2013) it can facilitate the creation of a burned area product with 10m resolution allowing improved post-fire evaluations on ecosystem damage and carbon emission.

While threshold and traditional machine learning based methods have successfully been applied to the task, they implicate drawbacks including the involvement of complex rule sets and requirement of feature engineering such as the extraction or generation of features from the raw data. In contrast, deep learning offers an end-to-end solution for image analysis and semantic segmentation. Among deep neural networks, convolutional neural networks (CNNs) show the most promising potential as they are designed to handle spatially dependent data such as images especially well (LeCun et al., 1998). Despite that, their share in solutions for burned area applications is comparatively low, which might be based on the limited amount of labelled training examples that is available.

Examples for optical data implementations include the CNNs developed by Pinto et al. (2019) and Pinto et al. (2020). Both are based on Visible Infrared Imaging Radiometer Suite (VIIRS) composites and active fire data and used to segment burned areas in selected locations. The framework proposed by Mohla et al. (2020) utilizes a variation of U-Net trained on weakly labelled Landsat data for burn scar identification in the Amazon rain forest. Park and Lee (2018) employ U-Net to map forest disasters including fires in Sentinel-2 imagery from South Korea. Using a combination of Sentinel-1 and Sentinel-2 images Lee et al. (2020) developed a model based on U-Net for bushfire detection in Australia. Other small studies also show promising results on the combination of U-Net and Sentinel-2 data for burned area

---

* Corresponding author

mapping in a mono-temporal approach. Beside Knopp et al. (2020), who achieved high overall accuracies in selected locations across the globe, Farasin et al. (2020) also proved the superior performance of this combination of network and sensor over conventional methods in a global case study of 147 cloud free areas. While these existing models are either trained and applied in a local or a global setting, this study investigates both by looking at the generalizability of a global model in comparison to local models using the example of two environmentally different research areas - Indonesia and Central Africa (Chad and Central African Republic).

In summary, one central part of this research is to transfer the promising combination of Sentinel-2 imagery and the neural network architecture U-Net to the present study areas. As fitting a neural network is based on a training with appropriate example data, the task of creating sample images is crucial. The underlying objective is consequently, to construct a high quality set of example images for training, validating and testing of the model in the defined research areas. The second goal is to compare two local (trained only with data from the respective research areas) and one global variant (trained with the whole dataset) of the model to determine the feasibility of using multiple local datasets within one common detection framework without implying a loss in accuracy.

2. MATERIALS AND METHODS

2.1 Definition of Sampling Areas

For defining the temporal and spatial distribution of sample data within the study areas, MODIS hotspots are used. This dataset contains point features that are generated by an algorithm within each pixels of size 1km that is flagged as containing one or more fires based on thermal anomalies (National Aeronautics Space Administration, 2020). The consideration of this dataset helps to refine the search for suitable, in this case defined as containing the target class, Sentinel-2 tiles that samples can be extracted from. Tiles are collected as evenly as possible across the study areas. In Indonesia, 21 Sentinel-2 images are used for the creation of the training, validation and test data, while in Central Africa 23 are selected. Additionally, two representative scenes with tile numbers not included in the reference datasets are selected for the use case simulation in each area. Overall, the dataset is a mix of cloud-free and cloudy images.

2.2 Data Preprocessing

Sentinel-2 tiles are downloaded as Level-1C files and passed through a pipeline applying atmospheric correction, band stacking and spatial resampling. The final output is a Level-2A product with 10m resolution consisting of 10 out of the original 13 Sentinel-2 bands (2, 3, 4, 5, 6, 7, 8, 8A, 11, 12). After their download, the selected scenes are classified manually scene-by-scene by domain experts to ensure high quality input data for the network. Both the original images and the created masks are then split into patches of 256x256px. Patches that would not include any burned area pixels are skipped. Using the previously described method, 6656 sample pairs are generated in total. This includes 3022 patches located in Indonesia and 3634 patches from the Central African research area.
Ensuring a neural networks ability to generalize is a core part of deep learning. To evaluate this characteristic, it is recommended to split the sample data into three different subsets, a training set, a validation set and a test set. In doing so, a trade-off between the statistical power in estimating the model performance on unseen data and the power of fitting a better adjusted model has to be

considered (Korjus et al., 2016). For this project, a data split ratio of 60%, 30%, 10%, for training, validation and test datasets respectively, is chosen. Three characteristics are thereby considered as recommended by Ng and Katanforoosh (2020). First, patches from one image are only assigned to one of the three groups. Second, the sets have the same distribution, both in order not to over- or underestimate the predictive power of the network, and third, the three groups are reproducible to make tests comparable.

2.3 U-Net Architecture

U-Net is a network architecture originally developed for biomedical image segmentation by Ronneberger et al. (2015). However, having characteristics like model simplicity, easy trainability and suitability for small datasets it has achieved good results in other fields as well (Zhang et al., 2020). It is a type of Fully Convolutional Network (FCN) only containing convolutional and no dense or fully connected layers, which means that it can accept images of any size as decisions are made on local scale as opposed to the global image. It has an end-to-end setting that allows the input of a raw image for producing a complete segmentation map as the output. The encoder-decoder architecture with two sequential paths compose its eponymous U-shaped form. The first path is a contracting path. It follows the typical architecture of a CNN consisting of convolutional and pooling layers and has the purpose of capturing the context and more complex features in the input image. The second path enables the localization of these features by using concatenations to combine the contextual information from the first path with the up-sampled output from the symmetric second part. The vector is thus retransformed to a two-dimensional image through transposed convolutions. Figure 1 shows a graphic representation of the utilized network.
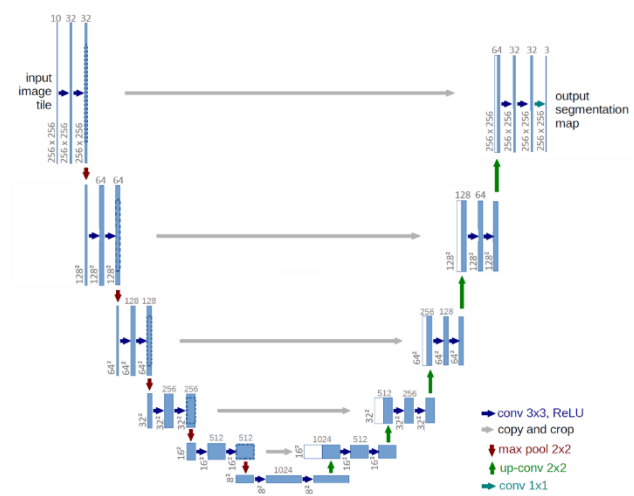

Figure 1. Architecture of the U-Net variation used in this study. Image source: Modified after Ronneberger et al. (2015)

2.4 Hyperparameter Settings

The implementation of the model follows the original architecture but includes some adaptions to account for differences in the input data and the related task. The model is set up with hyperparameters chosen according to best practice recommendations from developer platforms/communities as well as scientific papers or specified after some preliminary testing.

Batch Size

In comparison to the original paper, the batch size is increased to account for smaller input patches as well as larger variations within the samples. Preliminary testing showed that smaller batch sizes lead to large fluctuations in loss and accuracy during training. To counteract this effect, a batch size of 30 is defined. Samples are shuffled between epochs so that batches do not look alike, which ultimately makes the model more robust. While feeding samples into the algorithm all input images are standardized and the mask values are converted to one hot encoded rasters.

Weight Initialization

The initial setting of the network weights is determined using an activation-aware initialization. As ReLU (Rectified Linear Unit) is used as the activation function for the convolutional layers in this network, the Kaiming He initialization (He et al., 2015) is selected.

Regularization

The number of epochs the network is trained for is designed flexibly by setting the maximum number of epochs to 100 and including early stopping, which helps to avoid overfitting. Training is stopped when the loss on the validation set has not improved over more than 20 epochs. What is stored is the best model, here defined as the combination of weights that result in the lowest validation loss. Another way of avoiding overfitting is the use of dropout. A dropout layer with a dropout rate of 0.1 is therefore included after every convolutional block.

Optimization

For network optimization, Adam is used because of its ability to use adaptive learning rates for individual network parameters. With a learning rate of 0.001, β_1 value of 0.9 and β_2 value of 0.999, its parameters are kept at their default values. Additionally, the learning rate is reduced by the factor 0.5 when the validation loss does not improve over 10 epochs until a minimum of 0.0001 is reached. The additional learning rate adaption has been shown to improve convergence speed and accuracy (Smith, 2017).

Loss Function

As the case at hand presents a binary segmentation task, Binary Cross-Entropy (BCE) is the most suitable loss function for the proposed network and is used in the burned area model.

2.5 Experimental Setup

Network performance of the local and global variants is first evaluated on test data. This includes four scenarios: 1) A local classifier trained on the Indonesian training dataset and the resulting model evaluated in the Indonesian test data (I/I); 2) a local classifier trained on the Central African training dataset and the resulting model evaluated in the associated test data (CA/CA); 3) and 4) a global model trained on the whole dataset and separately applied to the Indonesian (G/I) and the Central African (G/CA) test data. This organization is based on the assumption, that in a real use case, the data the model is applied to would not consist of a mixture of patches from both regions but rather be confined to one. During the training phase, model performance is monitored on the training and validation set using loss, overall accuracy, and macro-averaged F1-score as metrics. Validation data is hereby used as the control to avoid over- or underfitting. Subsequently, the performance is reported on the test sets to compare the global and local classifiers. Therefore, predicted and true labels are contrasted in a confusion matrix to generate the number of True Positives (TP), False Positives (FP),

False Negatives (FN) and True Negatives (TN). As the model output consists of probability maps for each class, these are converted to binary masks using a 50% threshold before the comparison is made. Metrics derived from the confusion matrix are described in the following. Precision (P_x), or positive predictive value, denotes the number of successfully predicted samples that are in fact relevant. Recall (R_x) expresses the number of relevant samples that are successfully predicted.

$$P_x = \frac{TP}{TP + FP} \qquad (1)$$

$$R_x = \frac{TP}{TP + FN} \qquad (2)$$

The F1-score is defined as the harmonic mean of recall and precision metrics (Farasin et al., 2020). It describes the trade-off between the two previously explained metrics.

$$F_{1x} = 2 * \frac{P_x * R_x}{P_x + R_x} \qquad (3)$$

In the second phase, the three variants are additionally tested and evaluated on two example Sentinel-2 scenes per research area to mimic a real use case. The performance analysis of the network in a spatial context is set up to reveal strengths and shortcomings of the current training dataset which will allow for a targeted update and increase of the training data quality. The post-classification accuracy of the segmented scene pairs is analyzed using a set of random points, to compare the true classes with the classified data at the same locations. The sampling method used, is 'equalized stratified random', which in this case automatically distributes 100 points randomly within each class but generates the same number of points per class. The performance of each model variant is evaluated based on two scenes creating the same four scenarios as described before (I/I, CA/CA, G/I, G/CA). Point values are contrasted using a confusion matrix and the metrics described above.

The framework used for model implementation is Keras with a TensorFlow backend. Computing is conducted on an NVIDIATesla P100-PCIE with 27.4 GB of available RAM via Google Colab.

3. RESULTS

3.1 Performance Assessment based on test data

In order to determine whether one combined or two separate local classifiers are better suited for the task of burned area segmentation in Indonesia and Central Africa, the results of the global and local models are contrasted on the isolated test datasets in the following.

Model	Type	Metric	Burned	Non-burned
global	class	Precision (%)	88	96
		Recall (%)	83	97
		F$_1$ (%)	86	96
	overall	Accuracy (%)	94	
		Macro-F$_1$ (%)	91	
local	class	Precision (%)	83	96
		Recall (%)	91	97
		F$_1$ (%)	87	96
	overall	Accuracy (%)	94	
		Macro-F$_1$ (%)	92	

Table 1. Performance comparison of the **global** and the **local** model in **Indonesia** with predictions based on the **Indonesian** test dataset.

Table 1 shows the comparison of the local Indonesian and the global model. The accuracy assessment of the two variants applied to the Indonesian test data reveals a minor superiority of the local model (macro-F1: 92%) over the global model (macro-F1: 91%) with no difference in the OA at 94%. The results for the comparison between the local Central African and the global model applied to the same test data are displayed in Table 2. The accuracy of the global and local model is the same in both metrics (OA: 84%, macro-F1: 82%).

Model	Type	Metric	Burned	Non-burned
global	class	Precision (%)	70	91
		Recall (%)	78	87
		F_1 (%)	74	89
	overall	Accuracy (%)	84	
		Macro-F_1 (%)	82	
local	class	Precision (%)	69	91
		Recall (%)	80	86
		F_1 (%)	74	89
	overall	Accuracy (%)	84	
		Macro-F_1 (%)	82	

Table 2. Performance comparison of the **global** and the **local** model in **Central Africa** with predictions based on the **Central African** test dataset.

Overall, it can be observed that the accuracy for segmenting non-burned areas is higher than for burned areas. Precision and recall values additionally show that false predictions in the class burned are rather based on samples that are predicted as burned but are not relevant. Accordingly, shortcomings in the class non-burned can rather be attributed to relevant samples that are missed.

3.2 Use Case Accuracy Assessment

Consistent with the approach outlined in chapter 2.5, the local and global models' performance is compared based on entire scenes in the following.

3.2.1 Indonesia: The Sentinel-2 scenes chosen for the use case accuracy assessment in Indonesia have the tile numbers 49MCV and 48MTB. The former was acquired in West Kalimantan (2019-09-04) and the latter in Sumatra (2019-09-13).

Local Model

Class / GT	Non-Burned	Burned	Total	Recall (%)
Non-Burned	**100**	38	138	72.5
Burned	0	**62**	62	100.0
Total	100	100	**200**	
Precision (%)	100.0	62.0		**81.0%**

Table 3. Confusion Matrix of ground truth (GT) and classification (Class) for the use case example in **Indonesia** using the **local** model. The bottom right cell displays the overall accuracy. The F_1-score for non-burned is 84.7% and for burned 76.5%.

In contrast to the results displayed in the previous chapter the difference in accuracy between the two model variants is more visible (Table 3 and Table 4). The results agree in the higher accuracy of the non-burned class. Precision and recall as well as

the numbers within the confusion matrix again show that the lower accuracy in the burned class can be attributed to the detection of false positives, rather than false negatives. Accordingly, these falsely predicted samples appear as false negatives for the non-burned class resulting in precision values that are higher than the recall. As seen before, the accuracy measured by the F_1-score is lower for burned areas in comparison to non-burned areas for both models.

Global Model

Class / GT	Non-Burned	Burned	Total	Recall (%)
Non-Burned	**100**	44	144	69.4
Burned	0	**56**	56	100.0
Total	100	100	**200**	
Precision (%)	100.0	56.0		**78.0%**

Table 4. Confusion Matrix of ground truth (GT) and classification (Class) for the use case example in **Indonesia** using the **global** model. The bottom right cell displays the overall accuracy. The F_1-score for non-burned is 81.9% and for burned 71.8%.

True/false positive/negative predictions for burned areas
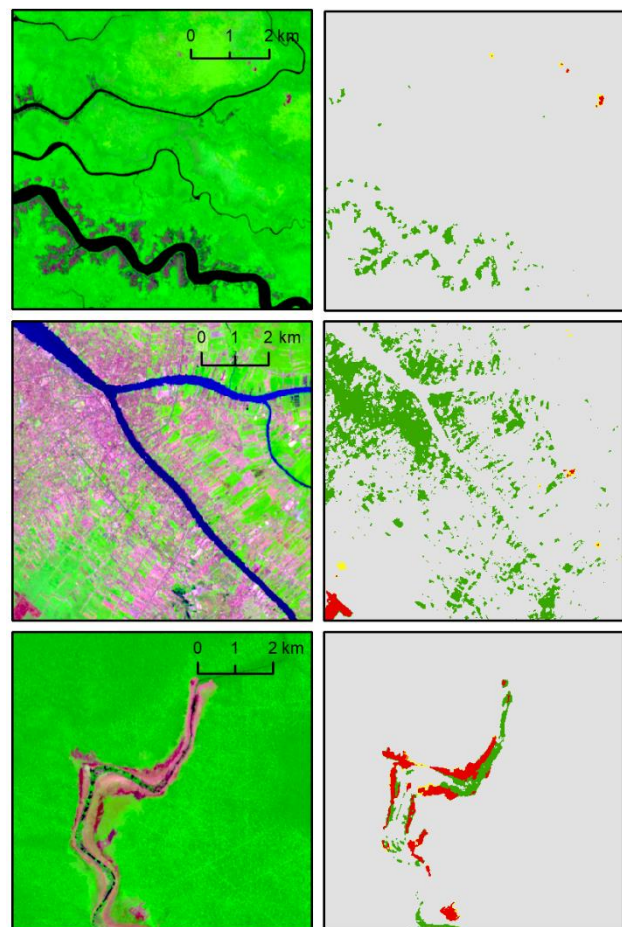
TN   FN
FP   TP

Figure 2. Example of false positive predictions in land cover classes like mangroves (top), urban (middle) and bare soil (bottom) in Indonesia (RGB composition: B12/B8/B3).

Analyses of the spatial results show that misclassifications of burned areas occur mainly in the land cover classes urban and bare soil but also in areas where mangroves are present (see Figure 2). Although confusions with cloud shadows are common in other studies (Chuvieco et al., 2019), the network performed well in this case (see Figure 3, top). Figure 3 also shows that in addition to large burns, small burn scars are segmented reliably.
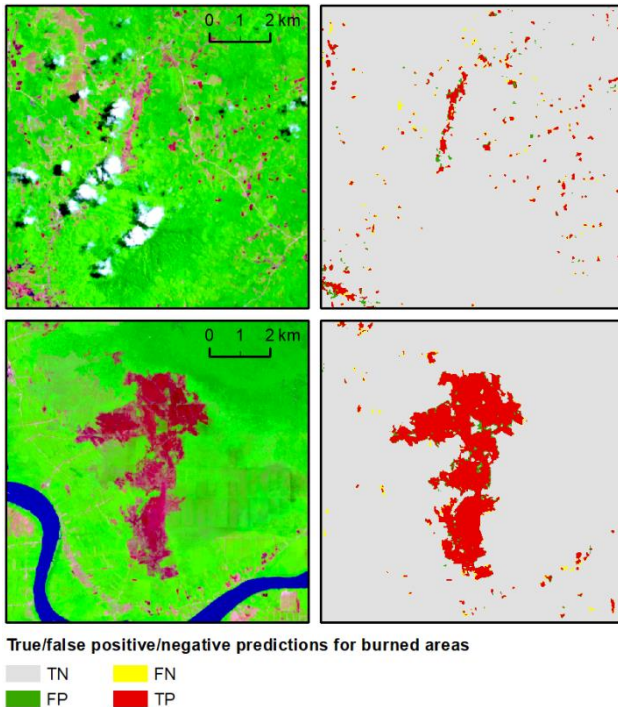

Figure 3. Segmentation of small and large burned areas in Indonesia (RGB composition: B12/B8/B3).

3.2.2 Central Africa: The Sentinel-2 scenes for the Central African research area are both located in Chad. The first image has the number 34PDU and was acquired on 2019-11-09, the second, with tile number 33PYN, was acquired on 2019-01-31.

The test results displayed in Table 5 and Table 6 confirm previous findings in terms of precision and recall. Similar to the Indonesian use case example, there is a small difference in the overall accuracy. In this case, the spatial analysis additionally showed that misclassifications can mainly be found in the land cover class bare soil. Clouds and cloud shadows are correctly labeled as non-burned (see Figure 4). The segmentation shown in Figure 5 represents an example of a successful prediction of true positives in the burned class.

Local Model

Class \ GT	Non-Burned	Burned	Total	Recall (%)
Non-Burned	**97**	21	118	82.2
Burned	3	**79**	82	96.3
Total	100	100	**200**	
Precision (%)	97.0	79.0		**88.0%**

Table 5. Confusion Matrix of ground truth (GT) and classification (Class) for the use case example in **Central Africa** using the **local** model. The bottom right cell displays the overall accuracy. The F_1-score for non-burned is 87.4% and for burned 84.5%.

Global Model

Class \ GT	Non-Burned	Burned	Total	Recall (%)
Non-Burned	**98**	16	114	86.0
Burned	2	**84**	86	97.7
Total	100	100	**200**	
Precision (%)	98.0	84.0		**91.0%**

Table 6. Confusion Matrix of ground truth (GT) and classification (Class) for the use case example in **Central Africa** using the **global** model. The bottom right cell displays the overall accuracy. The F_1-score for non-burned is 91.6% and for burned 90.3%.
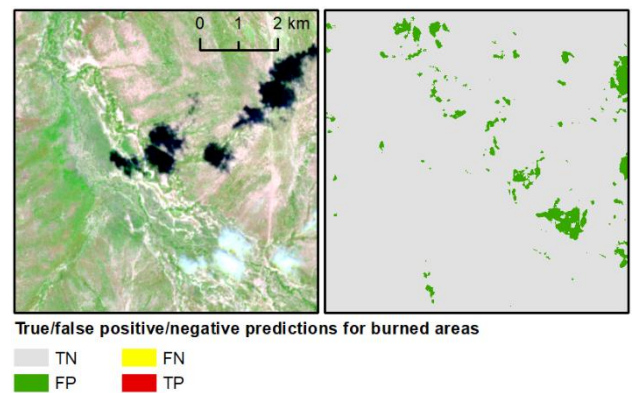

Figure 4. Example of confusions with dry and rocky terrain and reliable classification of cloud shadow as non-burned in the use case example in Central Africa (RGB composition: B12/B8/B3).
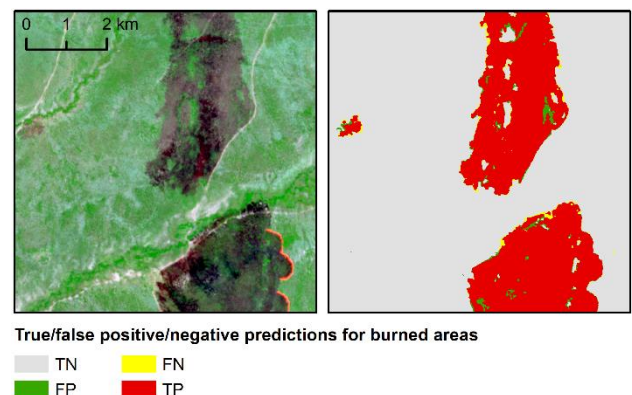

Figure 5. Segmentation of burned areas in Central Africa (RGB composition: B12/B8/B3).

4. DISCUSSION AND FUTURE RESEARCH

As observed in the spatial results, depending on the prevalence of different land cover types the performance is negatively affected. Misclassifications are observed for the burned area class when mangroves, urban areas or bare soil exist in the scene. These types of unburned areas have also caused problems in other segmentation approaches (Bettinger et al.). Figure 6 shows example spectra of these land cover types in reference to a burned area spectrum. The plot displays a high similarity of the spectral responses between classes.
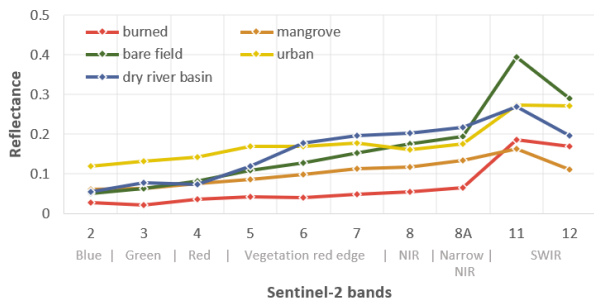
Figure 6. Example spectral signatures of burned areas and land cover types prone to misclassifications.

As the sampling design is targeted towards burned areas to reduce class imbalance, and fires mostly occur in forested regions or grassland, most of the negative examples in the training dataset represent vegetation, which has a spectral signature that is very different from the presented examples (see Figure 7). This explains the network's assignment of the affected pixels to the class with the most similar spectral reflectance, which in this case is burned.
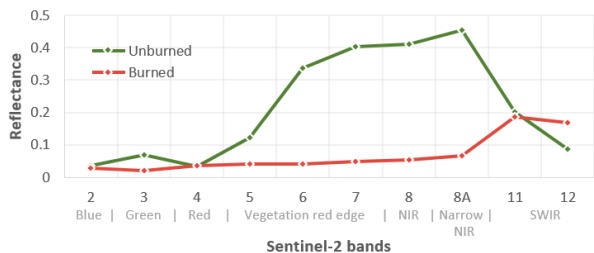

Figure 7. Example spectral signatures of unburned vegetation in comparison to burned area.

To account for these shortcomings, the goal in the next step of this project is to put together an improved training dataset. The non-burned area samples should be more heterogeneous so the model can cope with the more ambiguous differentiations. This will enable an updated comparison of class accuracies and is expected to lead to more similar results between the burned and non-burned class. While scale and image constitution could be factors, we would additionally like to investigate the origin of the divergence of global and local model performances in the use case examples further.

Other areas of future research include the implementation of k-fold cross validation for the performance assessment of different model configurations to make it more robust against other random factors like weight initialization and shuffling of training data between epochs (Brownlee, 2017). Furthermore, the evaluation of this model against a benchmark would aid the assessment of the results.

5. CONCLUSION

In this study, a model for burned area recognition in mono-temporal Sentinel-2 imagery from Indonesia and Central Africa was successfully developed. The established method uses a CNN for supervised segmentation of the target class that is based on U-Net, a network architecture specialized for segmentation problems with small training datasets. The neural network is trained from scratch with a customized training dataset for the research areas Indonesia and Central Africa. As the independent test identified some shortcomings in specific land cover classes

the main goal in the continuation of this project is the design of an updated catalogue of training samples to support the successful differentiation of burned areas and the identified land cover types prone to confusions with it. Overall, the accuracy assessments demonstrate state-of-the-art performance of the developed model as well as the effectiveness of this combination of network and datasets for the designated regions. The application of the models on test data shows the global model's ability to generalize despite the high dissimilarities between the two research areas, which proves the feasibility of using multiple local datasets within one common detection framework without implying a loss in accuracy.

REFERENCES

Bettinger, M., Martinis, S., & Plank, S., 2017: An Automatic Process Chain for Detecting Burn Scars Using Sentinel-2 Data. In *EARSeL 11th 2017*.

Brownlee, J., 2017: What is the Difference Between Test and Validation Datasets? *Machine Learning Mastery*. URL https://machinelearningmastery.com/difference-test-validation-datasets/ (accessed 04/20/2021).

Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M. P., Vilar, L., Martínez, J., Martín, S., Ibarra, P., La Riva, J. de, Baeza, J., Rodríguez, F., Molina, J. R., Herrera, M. A., & Zamora, R., 2010: Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling*, *221*(1), 46–58. doi.org/10.1016/j.ecolmodel.2008.11.017.

Chuvieco, E., Mouillot, F., Werf, G., San, J., Tanase, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker, T., & Giglio, L., 2019: Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sensing of Environment*, *225*, 45–64. doi.org/10.1016/j.rse.2019.02.013.

Farasin, A., Colomba, L., Palomba, G., Nini, G., & Rossi, C., 2020: Supervised Burned Areas Delineation by Means of Sentinel-2 Imagery and Convolutional Neural Networks. In *ISCRAM 2020 Conference Proceedings - 17th International Conference on Information Systems for Crisis Response And Management*. 1060-1071.

Kelly, L., & Brotons, L., 2017: Using fire to promote biodiversity. *Science*, *355*, 1264–1265. doi.org/10.1126/science.aam7672.

Knopp, L., Wieland, M., Rättich, M., & Martinis, S., 2020: A Deep Learning Approach for Burned Area Segmentation with Sentinel-2 Data. *Remote Sensing*, *12*, 2422. doi.org/10.3390/rs12152422.

Knorr, W., Jiang, L., & Arneth, A., 2016: Climate, CO_2 and human population impacts on global wildfire emissions. *Biogeosciences*, *13*(1), 267–282. doi.org/10.5194/bg-13-267-2016.

Korjus, K., Hebart, M., & Vicente, R., 2016: An Efficient Data Partitioning to Improve Classification Performance While Keeping Parameters Interpretable. *PLoS ONE*, *11*, e0161788. doi.org/10.1371/journal.pone.0161788.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., 1998: Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, 86 (11), 2278-2324, doi.org/ 10.1109/5.726791.

Lee, I. K., Trinder, J. C., & Sowmya, A., 2020: Application of U-Net Convolutional Neural Network to Bushfire Monitoring in Australia with Sentinel-1/-2 Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43B1*, 573–578. doi.org/10.5194/isprs-archives-XLIII-B1-2020-573-2020.

Lewis, S. L., Edwards, D. P., & Galbraith, D., 2015: Increasing human dominance of tropical forests. *Science*, *349*(6250), 827–832. doi.org/10.1126/science.aaa9932.

Mohla, S., Mohla, S., Guha, A., & Banerjee, B., 2020: Multimodal Noisy Segmentation based fragmented burn scars identification in Amazon Rainforest, URL https://arxiv.org/abs/2009.04634 (accessed 04/20/2021).

National Aeronautics Space Administration., 2020: *Fire Information for Resource Management System (FIRMS) | Earthdata: MODIS Active Fire Products*. URL https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms (accessed 04/20/2021).

Ng, A., & Katanforoosh, K., 2020: *CS230 Splitting into train, dev and test sets (Course Notes)*. URL https://cs230.stanford.edu/blog/split/ (accessed 04/20/2021).

Park, S. W., & Lee, Y., 2018: Detection of forest disaster using high-resolution satellite images with CNN (convolutional neural networks). *Journal of the Association of Korean Photo-Geographers*. 28. 87-101. doi.org/10.35149/jakpg.2018.28.4.006.

Pereira, A., Pereira, J., Libonati, R., Oom, D., Setzer, A., Morelli, F., Machado-Silva, F., Carvalho, L., 2017: Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires. *Remote Sensing*, *9*(11). doi.org/10.3390/rs9111161.

Pereira, J. M. (2003). Remote sensing of burned areas in tropical savannas. *International Journal of Wildland Fire*, *12*, 259–270. doi.org/10.1071/WF03028.

Pinto, M. M., Hurduc, A., DaCamara, C. C., Trigo, I. F., & Trigo, R. M. (2019). Burned Areas Segmentation with Convolutional Neural Networks. In *EGU General Assembly Conference Abstracts*.

Pinto, M. M., Libonati, R., Trigo, R. M., Trigo, I. F., & DaCamara, C. C., 2020: A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 260–274. doi.org/10.1016/j.isprsjprs.2019.12.014.

Pleniou, M., & Koutsias, N., 2013: Sensitivity of spectral reflectance values to different burn and vegetation ratios: A multi-scale approach applied in a fire affected area. *ISPRS Journal of Photogrammetry and Remote Sensing*, *79*, 199–210. doi.org/10.1016/j.isprsjprs.2013.02.016.

Ronneberger, O., Fischer, P., & Brox, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. 234–241. doi.org/ 10.1007/978-3-319-24574-4_28.

Roos, C. I., Scott, A. C., Belcher, C. M., Chaloner, W. G., Aylen, J., Bird, R. B., Coughlan, M. R., Johnson, B. R., Johnston, F. H., McMorrow, J., Steelman, T., 2016: Living on a flammable planet: interdisciplinary, cross-scalar and varied cultural lessons, prospects and challenges. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1696), 20150469. doi.org/10.1098/rstb.2015.0469.

Sizer, N., Anderson, J., & Winer, J., 2014: *High-Resolution Satellites Help Monitor and Respond to Fires in Southeast Asia*. URL https://wri-indonesia.org/en/blog/high-resolution-satellites-help-monitor-and-respond-fires-southeast-asia (accessed 04/20/2021).

Smith, L. N., 2017: Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 464–472. doi.org/ 10.1109/WACV.2017.58.

van der Werf, G., Randerson, J., Giglio, L., Leeuwen, T., Chen, Y., Rogers, B., Mu, M., van Marle, M., Morton, D., Collatz, G., Yokelson, R., & Kasibhatla, P., 2017: Global fire emissions estimates during 1997-2016. *Earth System Science Data*, *9*, 697–720. doi.org/10.5194/essd-9-697-2017.

Vetrita, Y., & Cochrane, M. A., 2020: Fire Frequency and Related Land-Use and Land-Cover Changes in Indonesia's Peatlands. *Remote Sensing*, *12*(1). doi.org/10.3390/rs12010005

Zhang, J., Li, X., Li, L., Sun, P., Su, X., Hu, T., & Chen, F., 2020: Lightweight U-Net for cloud detection of visible and thermal infrared remote sensing images. *Optical and Quantum Electronics*, *52*(9), 397. doi.org/10.1007/s11082-020-02500-8