



HAL
open science

Post Wildfire Burnt-up Detection using Siamese UNet

Sagar Verma, Kavya Gupta

► **To cite this version:**

Sagar Verma, Kavya Gupta. Post Wildfire Burnt-up Detection using Siamese UNet. ECML PKDD 2023 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Sep 2023, Turin, Italy. hal-04225474

HAL Id: hal-04225474

<https://hal.science/hal-04225474>

Submitted on 2 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Post Wildfire Burnt-up Detection using Siamese UNet

Sagar Verma¹ * and Kavya Gupta¹ **

Granular AI

Abstract. In this article, we present an approach for detecting burnt area due to wild fire in Sentinel-2 images by leveraging the power of Siamese neural networks. By employing a Siamese network, we are able to efficiently encode the feature extraction process for pairs of images. This is achieved by utilizing two branches within the Siamese network, which capture and combine information at different resolutions to make predictions. The weights are shared between these two branches in siamese networks. This design allows to effectively analyze the changes between two remote sensing images, enabling precise identification of areas impacted by forest wildfires in the state of California as part of ChaBuD challenge thereby assisting local authorities in effectively monitoring the impacted regions and facilitating the restoration process. We experimented with various model architectures to train ChaBuD dataset and carefully evaluated the performance. Through rigorous testing and analysis, we have achieved promising results, ultimately obtaining a final private score (IoU) of 0.7495 on the hidden test dataset. The code is available at <https://github.com/kavyagupta/chabud>. We also deploy the final model as a point solution for anyone to use at <https://firemap.io>.

Keywords: change detection · remote sensing · siamese networks · high resolution satellite imagery.

1 Motivation

Wildfires, caused by natural phenomena or human activities, impose substantial financial burdens, amounting to billions annually. Wildfires devastated more than 10 million acres in the United States alone in 2020. A significant portion of the associated costs stems from prevention efforts. To ensure the effective allocation of these funds, it is crucial to have a comprehensive understanding of high-risk areas and identify necessary precautions and mitigation solutions.

In the field of Earth Observation, the significance of crisis response and disaster management cannot be overstated. The timely and continuous monitoring of areas affected by catastrophic events like earthquakes, floods, and forest fires heavily relies on the availability of pertinent data. This research proposes an

* Corresponding author: sagar@granular.ai

** kavya.gupta100@gmail.com

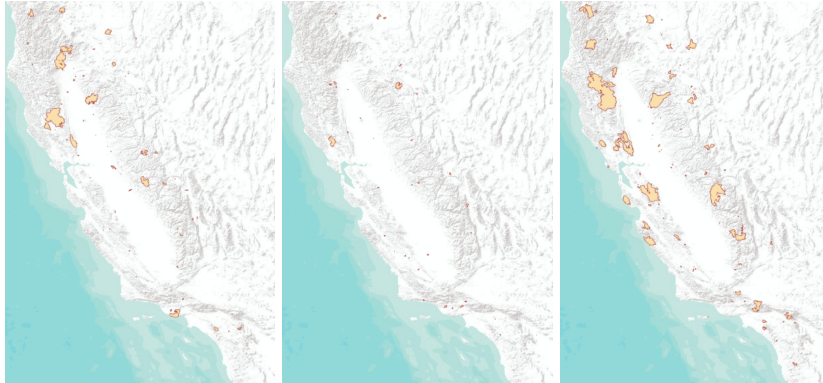


Fig. 1. Regions in California impacted by wildfires in the years 2018, 2019, and 2020.

advanced deep-learning approach for analyzing bi-date Sentinel-2 images to segment burnt-up areas after fires. The specific focus of this study revolves around leveraging the ChaBuD Dataset [10], which was released for the ECML-PKDD Competition. The ChaBuD challenge entails a computer vision task that uses raster bi-temporal geospatial data obtained from Sentinel-2 L2A satellite imagery. The primary objective is to detect regions in California that forest wildfires have previously impacted. Figure 1 shows wildfire-affected California regions in 2018, 2019, and 2020. This identification process is intended to aid local authorities in effectively monitoring the affected areas and facilitating the planning of restoration efforts.

The remainder of this paper is structured as follows. Section 2 provides an overview of the deep learning methods commonly employed in the literature for change detection tasks. Section 3 presents the details of our final model architecture, which achieved the highest private score in our experiments. We delve into the design choices, network configurations, and specific techniques employed to enhance the performance of our model. Section 4 is dedicated to the dataset used in our study and the experimental setup. Additionally, we present the experimental results, both quantitative and qualitative, to showcase the effectiveness of our proposed approach. These results are analyzed and compared to other models, demonstrating the superior performance of our method. Finally, Section 5 presents a comprehensive conclusion based on our findings.

2 State-of-the-art

Change detection is a pivotal concern within the remote sensing community as it effectively monitors the Earth on a global scale. By analyzing and modeling the changes in land usage and land cover (LULC) over time, we can understand the Earth's dynamics influenced by natural phenomena and human activities. The application of change detection techniques provides valuable insights into

various aspects of human intervention on Earth. For instance, it enables us to monitor urban sprawl, assess the levels of water and air contamination, identify illegal constructions, and more. Such information is crucial in developing a thorough understanding of the current state of the Earth’s surface and anticipating future LULC trends. Change detection techniques offer a powerful tool for researchers and policymakers to monitor and manage the Earth’s resources, promote sustainable development, and make informed decisions for the well-being of our planet. However, even if nowadays we can have access to a large amount of multi-temporal datasets [15] provided by satellites such as WorldView, Landsat, and Sentinel, the problem of change detection is very challenging. Traditional methods, summarised in surveys such as [9], use handcrafted techniques which heavily rely on pre-processing and post-processing making them not so easily adaptive to images that cover large areas. Change detection is a non-trivial problem as the accuracy of a method is highly influenced by registration errors [13] and illumination changes that do not really correspond to semantic changes.

With the remarkable advancements in deep learning across various domains, a wide range of change detection techniques have emerged. One notable approach, presented in [6], introduces a deep patch-based architecture. This method processes bi-temporal patches in parallel, leveraging a series of dilated convolutional layers to extract features. These features are then passed through a recurrent sub-network to capture sequential information. Finally, fully-connected layers are utilized to generate a change prediction map. While patch-based techniques have shown promising results, they need more time-consuming computations as they require processing each pixel individually. To address this issue, Daudt et al. [4] recently proposed three fully-convolutional Siamese networks based on the UNet architecture [11]. Their work aims to overcome the computational inefficiency of patch-based methods while accurately detecting regions with changes. By utilizing the UNet architecture in a fully-convolutional manner, their approach achieves efficient and effective change detection without the need for per-pixel processing. This advancement significantly reduces the computational burden, allowing for faster and more practical analysis of remote sensing imagery.

In the paper by Papadomanolaki et al. [7], the authors adopt a straightforward yet practical approach by employing a UNet architecture to extract spatial features from multi-date inputs. Additionally, LSTM blocks are incorporated into the model to capture and learn the temporal change patterns present in the data. This spatial and temporal information combination enables accurate change detection in remote sensing imagery. On the other hand, Chen et al. [3] propose an enhanced architecture called Siamese UNet, which builds upon the UNet framework. In their work, they introduce an attention mechanism to improve the detection of changed objects further. This attention mechanism allows the model to focus on relevant regions and prioritize informative features, enhancing the overall performance of change detection algorithms.

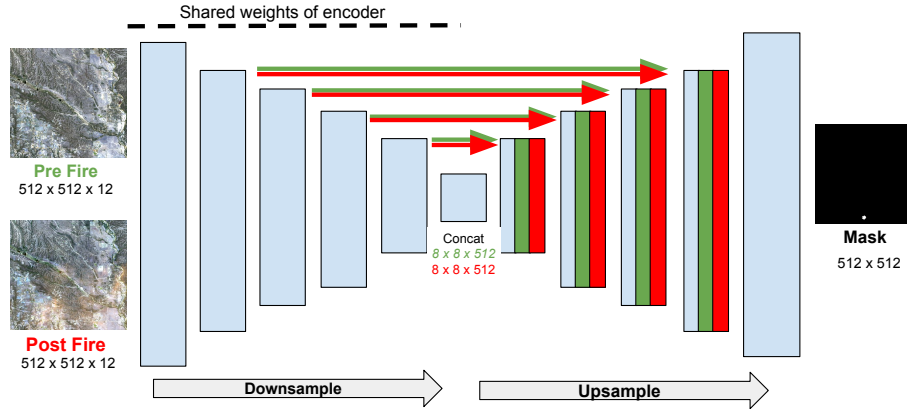


Fig. 2. Siamese UNet Concatenate architecture. The pre-fire and post-fire image are fed to a convolutional encoder with shared weights and concatenates the features of the paired images at difference resolutions.

3 Methodology

The employed network is based on the Siamese version of the UNet architecture, as depicted in Figure 2. Each block in the network comprises convolutional layers, batch normalization, and ReLU activations. The convolutional operations employ 3x3 filters with a stride and padding of 1. In the initial convolutional block, the depth of the output is increased to 64, while the height and width of the input volume remain unchanged. Subsequently, the following four blocks follow the same pattern, progressively doubling the depth and incorporating a 2x2 max pooling operation. Consequently, at the end of the encoder, the input undergoes downsampling, reducing its original dimensions to one-fourth and yielding 512 channels.

The encoding process is independently performed for each pre-fire and post-fire image. The features of the pre-fire and post-fire images after the encoding module are concatenated along the channels and fed to the decoder module. Subsequently, the decoder takes the last features from the encoder and performs upsampling to restore the original dimensions. Five convolutional blocks similar to the encoder are employed to achieve this, replacing max-pooling with 2x2 up-sampling operations. In addition, the resulting feature map from each upsampling process is concatenated with the corresponding features calculated from the symmetrical block in the encoder. This fusion of higher-resolution information with lower-resolution information enables the generation of more advanced features. Finally, at the end of the model, a 1x1 convolution operation is applied to compute the final probability heatmap, which detects areas of fire. This operation refines the features and produces a heatmap that indicates the likelihood of wildfires in different regions.

We also experimented with various variations of Siamese encoder-decoder networks, including UNet and DeepLab, using different encoding backbones such as ResNet-50/101 and MobileNetV3. Additionally, we explored the inclusion of auxiliary tasks such as input image pair reconstruction, in addition to burnt-up area segmentation. After conducting these experiments, we discovered that the Siamese UNet architecture with concatenation yielded the best performance.

3.1 Dataset and implementation details

All the experiments were conducted using the dataset provided as part of ChaBuD-ECML-PKDD [10], which consists of Sentinel-2A satellite images depicting forest fires in California. Twelve channels collecting information from the visible spectrum, infrared, and ultrablue are available for each image pair of pre-fire and post-fire images. For the challenge, we use the 1766 images to split into training and validation folds and the hidden test set to provide the final results (private leaderboard). We solely use the ChaBud Dataset and various data augmentations using Albumentations Library [1], with combinations of horizontal/vertical flips/random crops along with grid distortion, optical distortion, and elastic transform of the images.

Focal loss [5] is a highly effective loss function for training architectures, providing several advantages over traditional cross-entropy loss. Unlike cross-entropy loss, focal loss places more emphasis on challenging examples rather than solely focusing on confident predictions. Doing so ensures that the model continuously improves its predictions on hard samples, rather than becoming overly confident with easy ones. This approach effectively differentiates between easy and hard examples, enabling the training process to concentrate on learning from the most challenging samples within the dataset.

We use the Adam optimizer with momentum, with learning rate of 0.001 and momentum of 0.9. For focal loss, α is set at 0.99 and γ at 2.0. The batch size is set at 4 and the training is run for 400 epochs. All experimental setups were conducted using the PyTorch deep learning library [8] on a single NVIDIA RTX 3090 with 24 GB of GPU memory. We utilized GeoEngine platform [12,14] to track our experiments, and all the experiments configurations, weights, and metadata are available on GeoEngine ¹. The visuals results of the best and worst predictions are shown in Figure 3 and 4 respectively.

Architecture	Public Score	Private Score
Ours	0.7687	0.5436
Organizers	0.6765	0.5428

Table 1. IoU comparison results of Public and private score as on HuggingFace competition platform between ours and completion organizers.

¹ engine.granular.ai

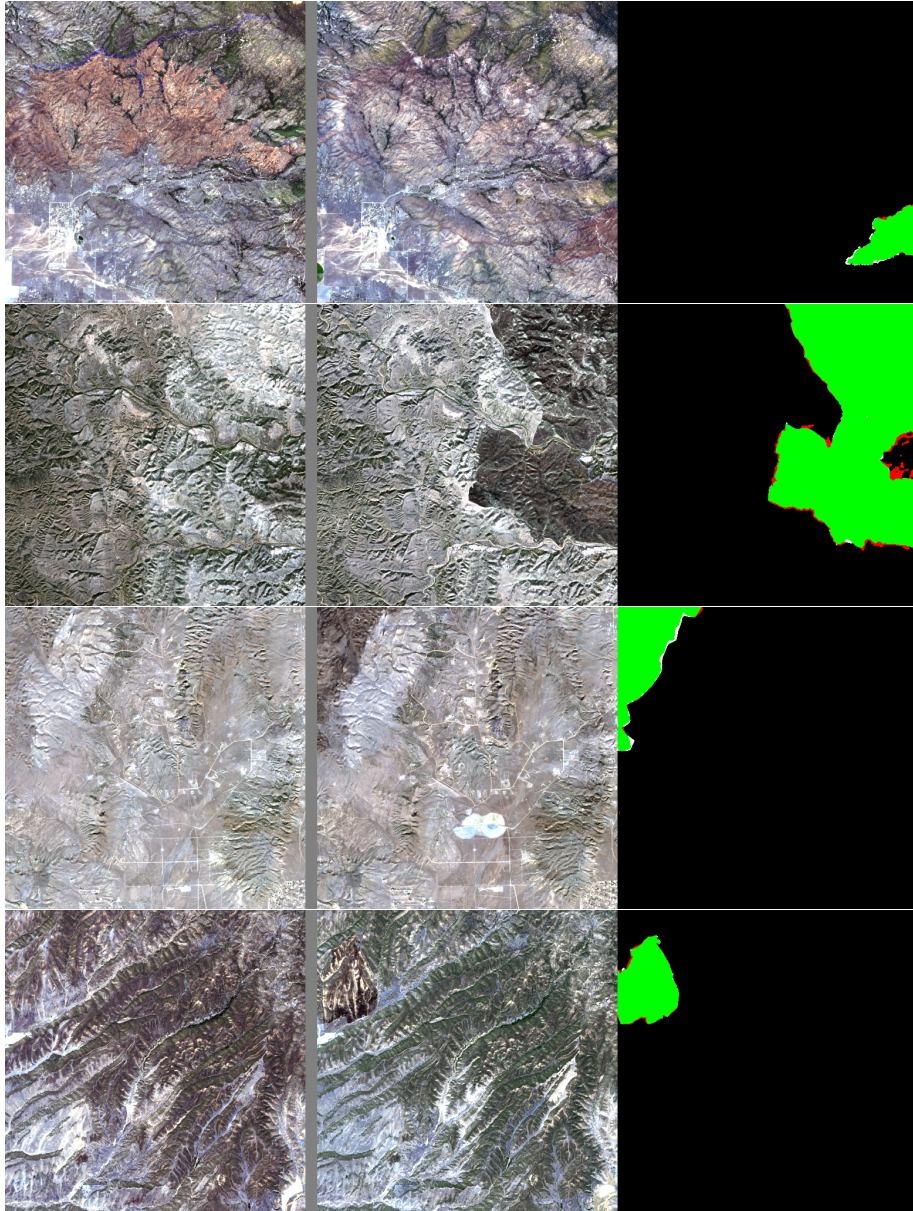


Fig. 3. The results display the best four predictions from the validation set. Each row showcases three columns: the first exhibits the pre-fire image, the second showcases the post-fire image, and the third presents the true and predicted masks. In the masks, white pixels represent the ground truth; green pixels indicate correct predictions, and red pixels denote false predictions.

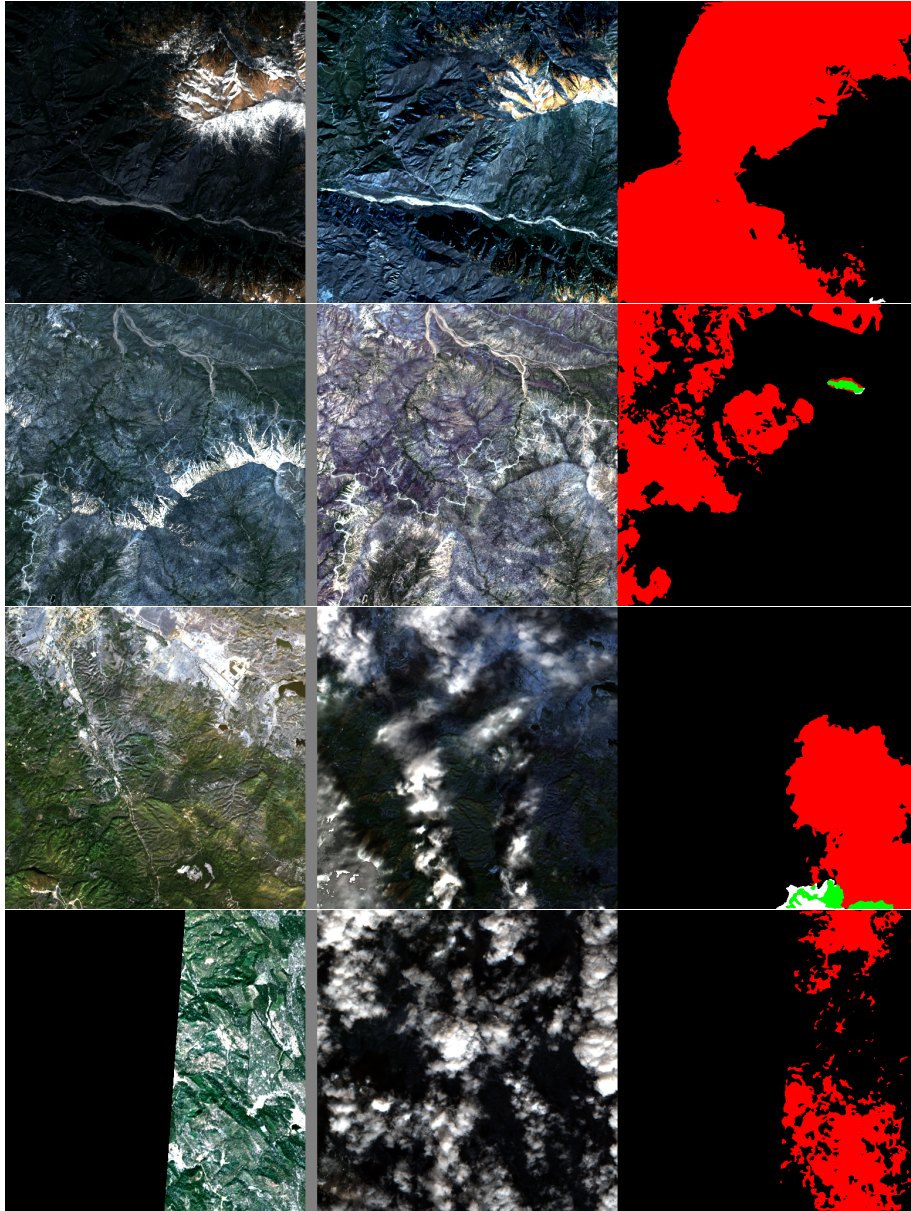


Fig. 4. The results display the worst four predictions from the validation set. Each row showcases three columns: the first exhibits the pre-fire image, the second showcases the post-fire image, and the third presents the true and predicted masks. In the masks, white pixels represent the ground truth; green pixels indicate correct predictions, and red pixels denote false predictions.

Architecture	Public Score	Private Score
DeepLab ResNet50 [2]	0.746	0.482
UNet Product [7]	0.755	0.492
Ensemble	0.768	0.543
DeepLab MobileNetV3 Large [2]	0.791	0.715
UNet Concatenate (Ours)	0.791	0.749

Table 2. IoU results comparison of Public Score (validation dataset) and Private Score (hidden test dataset). All architectures feature Siamese encoder heads with shared weights. For ResNet50 and MobileNetV3, we utilize pre-trained ImageNet weights, which require RGB channels scaled between 0 and 255 as input. The ensemble model combines Deeplab models trained on ResNet50/101/MobileNetV3 Large architectures and UNet models trained using Product/Concatenate techniques on various folds. The ensemble comprises a total of 20 models.

3.2 Conclusion

In this study, we introduced a Siamese version of the UNet architecture specifically designed to identify forest fire masks using the ChaBuD dataset derived from Sentinel-2 imagery. Our model achieved commendable performance on the change detection task, yielding accurate predictions of fire-affected regions. However, it is worth noting that our model did not incorporate temporal information during the training process. Accessing and leveraging temporal information could enhance the accuracy and reliability of fire mask predictions. By incorporating an RNN component into our Siamese UNet architecture, we can effectively capture the temporal changes and dependencies between different time steps, providing a more comprehensive understanding of fire dynamics. By leveraging the power of temporal information and incorporating it into the training process, we anticipate significant improvements in fire mask prediction accuracy. This enhancement would enable our model to better adapt to changing fire patterns, resulting in a more precise delineation of fire-affected areas. Future research endeavors should focus on incorporating temporal information and exploring innovative approaches to leverage the dynamic nature of multi-temporal satellite imagery for improved forest fire detection and monitoring.

References

1. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. *Information* **11**(2) (2020)
2. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017)
3. Chen, T., Lu, Z., Yang, Y., Zhang, Y., Du, B., Plaza, A.: A siamese network based u-net for change detection in high resolution remote sensing images. *IEEE JSTARS* **15**, 2357–2369 (2022)
4. Daudt, R.C., Saux, B.L., Boulch, A.: Fully convolutional siamese networks for change detection. In: *ICIP* (2018)

5. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
6. Mou, L., Bruzzone, L., Xiang, Z., Zhu, X.: Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *CoRR* **abs/1803.02642** (2018)
7. Papadomanolaki, M., Verma, S., Vakalopoulou, M., Gupta, S., Karantzalos, K.: Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data. In: IGARSS. pp. 214–217 (2019)
8. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
9. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image change detection algorithms: a systematic survey. *IEEE TIP* **14**(3) (2005)
10. Rege Cambrin, D., Colomba, L., Garza, P.: Cabuar: California burned areas dataset for delineation. *IEEE Geoscience and Remote Sensing Magazine* (2023). <https://doi.org/10.1109/MGRS.2023.3292467>
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
12. Shin, H., Exe, N., Dutta, U., Joshi, T.R., Verma, S., Gupta, S.: Europa: Increasing accessibility of geospatial datasets. In: IGARSS (2022)
13. Vakalopoulou, M., Karantzalos, K., Komodakis, N., Paragios, N.: Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data. In: CVPRW (2015)
14. Verma, S., Gupta, S., Shin, H., Panigrahi, A., Goswami, S., Pardeshi, S., Exe, N., Dutta, U., Joshi, T.R., Bhojwani, N.: GeoEngine: A platform for production-ready geospatial research. In: CVPRD. pp. 21416–21424 (2022)
15. Verma, S., Panigrahi, A., Gupta, S.: QFabric: Multi-task change detection dataset. In: CVPRW. pp. 1052–1061 (2021)