The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-4-2024
45th Canadian Symposium on Remote Sensing (CSRS), 10–13 June 2024, Halifax, Canada

Improving Off-Nadir Deep Learning-Based Change and Damage Detection through Radiometric Enhancement

Samuel McNally [1], Alia Nielsen [1], Aidan Barrieau [1], Shabnam Jabari [1]

[1] Department of Geodesy & Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada - (samuel.mcnally, alia.nielsen, aidan.barrieau, sh.jabari)@unb.ca

Keywords: Change Detection, Off-nadir Imagery, Remote Sensing, Neural Networks, Radiometric Enhancement.

Abstract

Aerial and satellite imagery can provide vital information to relief organizations about the extent and distribution of damages after natural disasters. With manual change detection being too inefficient to be effective, the pursuit of automated change detection has accelerated with the recent developments of deep learning methods. Off-nadir imagery (captured not directly overhead) is the fastest to acquire post-disaster, making it ideal for disaster management scenarios. However, the changes in viewing angles result in shadows and occlusions, making damage detection more difficult. Differences in illumination conditions are ever present in bitemporal aerial and satellite imagery, especially for off-nadir imagery, where the reflectance angle affects the amount of light returning to the sensor, making it harder to detect changes and damages. The hypothesis of this study was that artificial intelligence methods fail to adequately account for the illumination differences between images. To test this hypothesis, two radiometric enhancements, matching and equalization, were applied to four change and damage detection datasets, including a damage detection dataset from the 2010 Haiti earthquake. Using a leading high accuracy fusion convolutional neural network architecture called Changer, improvements of up to 20 percent for F1-Score, a popular remote sensing metric for quantifying the number of correctly classified pixels for specific datasets, were achieved through applying radiometric enhancement techniques. Applying radiometric enhancements on a case-by-case basis led to considerable improvements in accuracy, showing the promise of radiometric enhancement. Lower accuracies were achieved on the Haiti dataset, outlining the need for large disaster-specific datasets for training.

1. Introduction

Natural disasters like earthquakes have profound societal impacts, necessitating swift relief efforts. Quantifying how much damage and where the damage has occurred is vital for rescue, relief, and reconstruction efforts. Post-disaster imagery from satellites, planes, or Unmanned Aerial Vehicles (UAVs) can provide a timely way to assess damages after a natural disaster by looking at building rooftops. Manual damage detection is a time-consuming and inefficient task, prompting the use of automated methods to detect damages. The leading techniques to detect these changes use a form of artificial intelligence called convolutional neural networks. Novel network architectures are frequently being developed to improve the accuracy of building change and damage detection.

Neural networks necessitate large datasets for training so that the networks can learn what changed or damaged buildings look like and adjust the functions within their network to detect these features. The ability of a neural network to detect changes and damages is limited by its training data. Most current change and damage detection datasets are comprised of nadir-angle images, where the camera is directly overhead (Shen et al., 2021; Pang et al., 2023). While nadir imagery can yield better results, the delay in image acquisition makes it less practical than off-nadir imagery (Shen et al., 2021). The reported accuracies on nadir datasets are less relevant for disaster response purposes as models trained on nadir-angle images will not have the same performance on the quicker-to-acquire off-nadir-angle images.

There is no consistent angle value that determines when an image is “off-nadir”. For instance, the S2Looking dataset (Shen et al., 2021) discerns images as off-nadir when the viewing angle surpasses 15 degrees away from the nadir. Meanwhile, other sources (Wang et al., 2022) advocate for an angle of 25 degrees to classify an image as “off-nadir.” The farther the off-nadir angle for the images, the harder it is to detect change and damage for humans and neural networks alike. Some challenges with off-nadir images are illumination differences, perspective distortion, shadows, visible facades, occlusions, and shifts in rooftop locations (Jabari & Zhang, 2017, Wang et al., 2022, Pang et al., 2023). Differences in illumination conditions are ever present in bitemporal aerial and satellite imagery, especially for off-nadir imagery where the reflectance angle affects the amount of light returning to the sensor (Chen & Shi, 2020). In addition, differing atmospheric conditions, time of day, and time of year all contribute to changes in illumination conditions between pre- and post-disaster images (Chen & Shi, 2020, Jabari & Zhang, 2017). Due to the additional challenges with off-nadir imagery, current networks are still much less accurate on off-nadir imagery change detection tasks, leading to unsatisfactory results (Shen et al., 2021, Pang et al., 2023, Fang et al., 2022). Any improvements in accuracy are vital as they could lead to better rescue, relief, and reconstruction efforts when disasters strike.

Based on the available data, there appears to be a lack of damage detection datasets for training and testing neural networks. How networks trained on change detection perform on damage detection tasks has not been thoroughly tested.

1.1 Objectives

The first goal was to test the accuracy of a leading neural network on earthquake-specific imagery by creating a new dataset. The second goal was to improve the accuracy of deep learning methods on off-nadir imagery by addressing the differences in illumination conditions between images through radiometric enhancement.

By improving upon existing methods, this research aimed to enable swifter and more precise damage detection based on off-nadir images, with the ultimate goal of improving the way support could be delivered to those experiencing unimaginable loss due to disaster.

1.2 Dataset Selection

The quality of results from a network depends on the quality and quantity of the training dataset used. The training data should provide a representative sample to achieve the best accuracies across various testing scenarios. Thus, the choice of datasets was pivotal to this project. A wide range of building change and damage detection datasets from aerial and satellite imagery are publicly available (Shen et al., 2021). Three primary change models classify label annotations between image pairs so a network can learn how to identify change: a binary, scaled, or categorical change model. Binary models (Ji et al., 2018; Holail et al., 2023) describe binary changes (e.g. changed vs unchanged or damaged vs not damaged), whereas scaled models (Gupta et al., 2019, Abdi et al., 2022) provide more detailed classifications. In the case of Gupta et al., for their dataset xBD, they created a "joint damage scale" ranging from no damage (0) to completely destroyed (3) (Gupta et al., 2018). Categorical models (Ji et al., 2018) classify the pixels into semantic classes, such as grass, building, and water. Categorical and scaled models have different options, which can be more challenging than the binary annotations for networks to classify (Wang et al., 2022) and have less compatibility between datasets. Therefore, the xBD dataset was not selected for this project despite being specific to natural disasters. With the lack of datasets being a central issue for the change detection field, it becomes even more of an issue when looking at damage detection. For example, in comparing datasets for change detection and damage detection, many instances of change are not equivalent to damage like new buildings.

There is a lack of available datasets for network training and testing, particularly for off-nadir angle imagery, with BANDON and S2Looking being two of the few public off-nadir image datasets. The building change detection with off-nadir aerial images dataset (BANDON) was chosen for training due to its large size, urban setting, and compatibility with semantic and binary change detection (Pang et al., 2023). The BANDON dataset is composed of 2283 off-nadir 2048 × 2048 image pairs depicting 123,000 change instances (Pang et al., 2023). The BANDON dataset includes change masks and three semantic classes: building rooftops, facades, and background pixels (Pang et al., 2023). The binary labels classify all pixels as changed (like a new building or a damaged building) or unchanged (all other pixels). Additional datasets were used for testing: S2Looking (Shen et al., 2021), LEVIR-CD (Chen and Shi, 2020), and a Haiti dataset (Abdi et al., 2021). The S2Looking dataset is a large off-nadir satellite image dataset with changes and damages consisting of 5000 bitemporal off-nadir image pairs of rural areas with 65,920 annotated instances of change (Shen et al. 2021). It was used for testing instead of training since it did not have the additional semantic information and captured more rural areas, leading to an imbalance in the dataset with very few change pixels. Another dataset chosen for testing was the LEVIR-CD dataset, as it is one of the most popular change detection datasets, so it is useful for comparison. It consists of 637 high-resolution image pairs from Google Earth with a size of 1024 × 1024 pixels with a total of 31,333 instances of change. The final dataset used for testing was a dataset from the 2010 Haiti earthquake created by Abdi et al. (2021) that was modified for this project to test the network on earthquake-specific images with extensive damage.

Pre- and post-disaster multispectral training images are occasionally supplemented with additional data. Supplementary data sources include building footprints, hyperspectral imagery, radar imagery and coordinates for the images (Abdi et al., 2021; Ji et al., 2019; Li et al., 2023; J. Liu et al., 2018). Many change detection datasets like BANDON and S2Looking do not have these additional data sources, limiting the possible computations on the dataset. The Haiti dataset had one georeferenced image, allowing one image to be projected onto the other.

1.3 Network Selection

Detecting damaged or changed buildings from aerial and satellite imagery is crucial to providing timely and accurate information for natural disaster relief and reconstruction efforts. In recent years, several types of deep learning neural networks have been developed to automate this challenging task accurately and efficiently. The primary objective of these networks is to increase accuracy and reduce the training and testing time. Existing networks continue to have difficulty achieving high accuracies, especially when confronted with off-nadir images that can be acquired quickly after natural disasters (Pang et al., 2023, Shen et al., 2021).

The accuracy of building change detection networks is determined by several metrics, including IOU, F1, Recall, Precision, Overall Accuracy, and other custom-made metrics (Codegoni et al., 2022; Pang et al., 2023; Shen et al., 2021). These metrics are ratios between true positives, false positives, true negatives, and false negatives. They attempt to quantify the number of correctly identified pixels. The accuracy of the networks varies greatly depending on the dataset. Two of the most common metrics are F1-Score (or F-Score), the harmonic mean of precision and recall as shown in Equation 1, and Intersection Over Union (IOU), the number of overlapping pixels divided by the union of the two classes.

$$F1 - Score = \frac{2TP}{TP + 0.5(FP + FN)} \tag{1}$$

where TP = true positives
FP = false positives
FN = false negatives

Testing accuracies are the most pertinent metrics as they show how a network performs on unseen data, mirroring how disaster relief organizations could use these networks. Neural networks that are too complex run the risk of overfitting, causing the testing accuracy to be much lower than its training accuracy. Most existing neural networks for building change detection use supervised learning with labelled before and after training images. Existing networks are primarily built using PyTorch, a cutting-edge Python deep-learning library (MetaAI, 2023; Shen et al., 2021).

With deep learning methods constantly evolving, novel network architectures are being developed, optimized, and tested for detecting building changes in satellite imagery. The Convolutional Neural Network (CNN) architecture forms the basis of most change detection deep learning models, where trainable filters are passed over images to pick out features. U-net and Resnet are two of the most popular backbones for CNNs as they use skip connections to expedite training (Gupta et al., 2019; Ji et al., 2019).

Optimizing the architecture of a network and the preprocessing steps used in building change detection networks can improve results. An online dashboard that compares the accuracy of networks on a nadir satellite imagery dataset, LEVIR-CD, shows that the Changer network achieved the highest F1 and IOU scores to date, at 92.33 and 85.76 (MetaAI, 2023). When dealing with more complex datasets, notably ones with off-nadir imagery, existing networks have produced lower accuracies with F1 and

IOU testing accuracy scores in the 50 to 70 range (Pang et al., 2023; Shen et al., 2021).

The Changer network architecture developed by Feng et al. in 2023 was selected for this project primarily for its high accuracy on change detection datasets including S2Looking, BANDON, and LEVIR-CD as demonstrated on online dashboards and when comparing their research to others (Fang et al., 2022, MetaAI, 2023). The Changer network is a convolutional neural network with a ResNet backbone (Fang et al., 2022). It has feature interaction layers to exchange spatial and channel information to gain the context of pixels, and fusion layers to improve the spatial alignment between images (Fang et al., 2022). Cross-entropy loss and the AdamW optimizer are used as hyperparameters for the Changer network (Fang et al., 2022). The Changer network uses the mmsegmentation library, a subset of Pytorch, a common deep-learning Python library.

1.4 Preprocessing Steps

Preprocessing can be pivotal to improving the accuracy of networks. To increase the number of training samples obtained from one dataset, colour jittering, rotation, flips, and cropping can be applied to the training images (Fang et al., 2022; Pang et al., 2023). Downsampling of images can expedite training (T. Liu et al., 2021; Pang et al., 2023; Shen et al., 2021). These preprocessing steps do not solve the problem of illumination differences in bitemporal image pairs. Performing calculation preprocessing tasks on datasets can improve accuracy, yet it can introduce additional error sources (Pang et al., 2023; Shen et al., 2021). Radiometric enhancements have not been extensively applied to change detection datasets.

2. Methodology

2.1 Applying Radiometric Enhancements

To test whether radiometric enhancement can improve the final accuracy of change detection by addressing illumination differences between images, two types of radiometric enhancement, equalization and matching, were implemented on the four selected datasets: BANDON, S2Looking, LEVIR-CD, and the Haiti dataset. The hypothesis was that, by reducing the variations in the spectral characteristics of the images, the neural network could focus on detecting building changes instead of the changes in colour between images.

Histogram equalization evenly spreads the brightness values across the spectrum. The first step is to create the cumulative histogram for each band. Next, values in the cumulative histogram are multiplied by the number of brightness values – 1 and divided by the total number of pixels in the image. These new computed brightness values replace the old brightness values for every pixel, with the resultant equalized images having the same number of pixels across the entire spectrum. In the case of 8-bit images, the brightness values are evenly distributed from 0 to 255. Equalization was effective at highlighting dark features and making the spectral profiles identical, although it did cause colour distortion.

A second radiometric enhancement technique, histogram matching, was also applied, altering the colours of the after images to match the colours of the before images with less colour distortion than equalization. The after image was chosen to be matched to the before image since, in natural disaster scenarios, the before image is the baseline to compare against, so it was left unchanged. Firstly, cumulative histograms were made for each pair of before and after images. The brightness values in the cumulative histogram of the before image replaced the brightness values in the cumulative histogram of the after image with the closest frequency. As a result, the histograms of the after image matched the shape and average pixel value of the before images. Figure 1 shows an example of equalized and matched images, with the matched after image now having the same dominant green colour. For the BANDON dataset with three timestamps for some images, the second and third images were matched to the earliest timestamp that formed a pair of images. Despite changing the colours in the after image with matching and equalization, changes in the buildings were still easily visible between images.

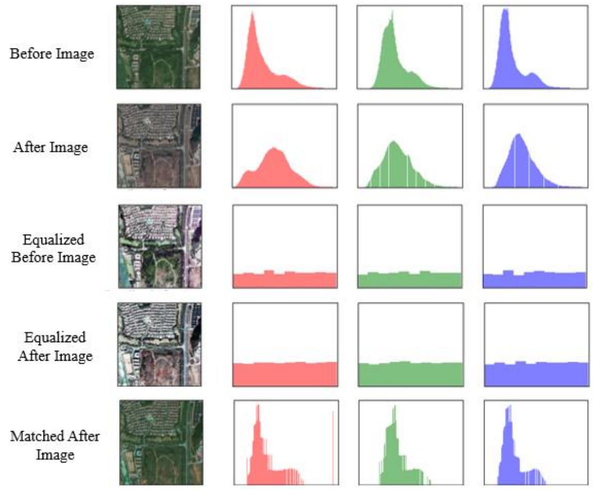

Figure 1. BANDON Radiometric Enhancement Example

A radiometric enhancement software was created in Python to efficiently apply equalization and matching to entire datasets at a time. It is available on GitHub for future use at https://github.com/SMProgrammer/RadiometricEnhancement.

2.2 Network Training and Testing

The training was performed in Python in Google Colab with a V100 Graphic Processing Unit and took approximately 8 hours for each network. The number of iterations used for training was 40,000 since it is a standard training time for this type of network (Fang et al., 2022; Pang et al., 2023), and the results were levelling off at that number of iterations. The testing generally took two to three minutes per dataset, with about a second required per image. The speed of the test results shows that a relief organization could use a network to obtain timely results in a real-world natural disaster scenario.

The same radiometric enhancement technique was applied to the training and the testing data. First, the Changer network was trained on the regular BANDON dataset and tested on the test portion of the BANDON dataset and the other three datasets to achieve baseline accuracies for comparison. Second, the Changer network was trained on the equalized BANDON training images and tested on the equalized datasets. Third, the Changer network was trained on the matched BANDON training images and tested on the matched datasets. These three steps were repeated with the semantic information included for the BANDON and Haiti datasets, resulting in a total of six networks being fully trained and tested. Semantic testing and training were only performed for the BANDON and Haiti datasets, as they were the only datasets compatible with semantic change detection. Configuration files were created for the BANDON and Haiti datasets to make them compatible with the Changer network for semantic and non-semantic change detection.

2.3 Earthquake Dataset Preparation

The Haiti dataset, developed by Abdi et al. (2021) was modified to be compatible with image-level damage detection for a network trained on the BANDON dataset. The original Haiti dataset contained a 2009 0.5-meter resolution off-nadir pre-disaster image, a 2010 post-disaster nadir 0.5-meter resolution georeferenced orthophoto, and a building footprint polygon layer.

Firstly, the pre-disaster image was georeferenced to the post-disaster orthophoto using the Auto-Georeference tool in ArcGIS Pro as it yielded superior results to manual georeferencing. The registration used a second-order polynomial transformation with 619 control points to achieve RMS forward, inverse and forward-inverse error values of 2.642, 2.662, and 0.015. Secondly, the building footprints were projected and snapped onto the 2010 orthophoto to create a raster while maintaining the same cell size, extent, and coordinate system. Thirdly, a change mask with only collapsed pixels, a raster with only non-collapsed buildings, and a raster with all building images was created. Having this semantic information with which pixels were buildings allowed the Haiti dataset to be compatible with the semantic classes in the BANDON dataset. Fourthly, the images, change mask, and building rasters were split into forty-one 512 by 512 images which corresponded to the crop size applied by the Changer network to ensure full compatibility and to maximize the number of image pairs the dataset could provide.

3. Results

The results indicated that histogram matching and equalization can lead to notable accuracy improvements in certain situations. The visual results for non-semantic testing on the S2Looking and Haiti datasets are shown in Figures 1 and 2. The image on the bottom left shows the ground truth labels with white pixels indicating changes or damages. The regular, matched and equalized predictions from the three non-semantic networks are shown in the rest of the bottom row.

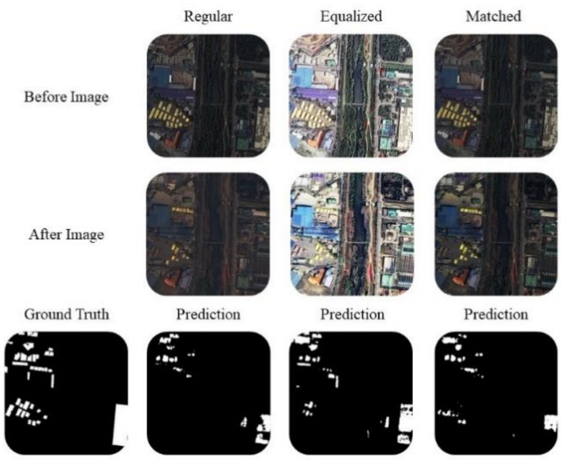

Figure 2. Visual Results for a Pair of S2Looking Images

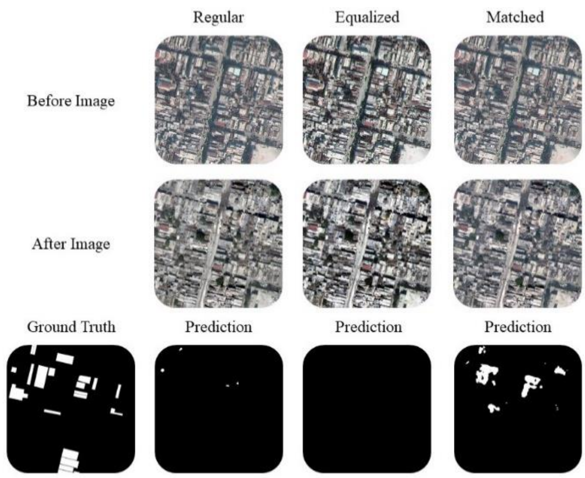

Figure 3. Visual Results for a Pair of Haiti Images

The F1-Score and IOU for the changed class and the mean for both the changed and unchanged classes for the non-semantic testing are shown in Table 1 and displayed in Figure 4. LEVIR-CD was the dataset with the highest accuracy scores. Depending on the dataset, different types of enhancements produced the highest scores.

Dataset	Enhancement	Changed Class		Mean Metrics	
		F1-Score	IOU	F1-Score	IOU
BANDON	Regular	66.03	49.29	82.55	**73.73**
	Equalized	**66.54**	**49.86**	81.96	73.06
	Matched	66.12	49.38	**82.57**	**73.73**
Haiti	Regular	7.82	4.07	52.27	48.86
	Equalized	4.77	2.44	50.73	48.01
	Matched	**28.49**	**16.61**	**62.55**	**55.03**
S2Looking	Regular	32.40	19.33	65.94	59.15
	Equalized	**50.53**	**33.81**	**74.27**	**65.73**
	Matched	15.14	8.19	57.28	53.53
LEVIR-CD	Regular	**70.07**	**53.93**	**84.16**	**75.25**
	Equalized	68.64	52.25	82.23	73.09
	Matched	62.07	45.00	80.07	70.61

Table 1. Summary of Non-Semantic Testing Accuracies

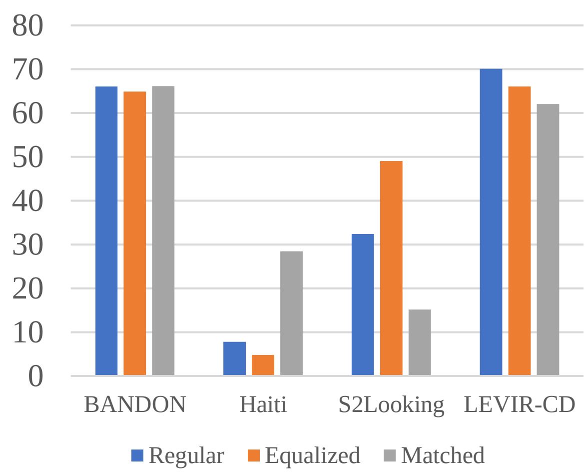

Figure 4. Numerical Testing Results - Changed Class F1-Score

The Haiti dataset saw the lowest accuracies of all the datasets tested. Applying histogram matching to the results for the non-semantic network produced an increase of over 20.7 in F1-Score on the changed class and a 10.3 increase in the mean F1-Score between the changed and unchanged class. Visually, the matched prediction picked up more damaged buildings and more pixels of damaged buildings, leading to higher scores (see Figure 3).

Histogram equalization resulted in major increases in accuracy for S2Looking. The equalization method improved the F1-Score by 18.1 and the IOU by 14.5 for the changed class. Applying radiometric enhancement techniques to the BANDON and LEVIR-CD datasets resulted in minor increases and decreases in accuracy.

In general, testing networks on data that had been enhanced differently than their training data led to lower accuracies. However, since the equalized S2Looking drastically improved results, networks trained on the regular and matched BANDON data were also tested on the equalized S2Looking images. The network trained on the regular S2Looking images saw an improvement of 10 in F1-Score when tested on equalized S2Looking images.

4. Discussion

Applying radiometric enhancements showed much promise as it led to substantial accuracy improvements for the S2Looking and Haiti datasets. The Haiti dataset provided a challenge for a network trained on change detection as that dataset contained exclusively damages, leading to lower accuracies.

There were much higher precision scores than recall scores for the changed class for all the datasets except for LEVIR-CD. In addition, the unchanged class scored highly at over 93% on all metrics for all networks and all datasets. These results indicate that the networks were not detecting enough changes or damages, but the changes or damages the networks did detect were generally correct. Even the Haiti dataset, with low scores overall, still had precisions over 50 percent for the changed class. To make the networks focus more on the changed pixels, weighting the loss function in future research may be necessary to address the imbalance in the number of pixels in the changed and unchanged classes.

The addition of the semantic information for the BANDON and Haiti datasets made little difference to the achieved accuracies, indicating that future datasets may not need this additional information. One case where additional semantic information may be useful is labelling what type of natural disaster has occurred in an image, as ideally, a network would affect its pixel classifications.

4.1 BANDON Validation Accuracies

There is inherent variability in training AI models. The networks were trained for 4,000 iterations several times, resulting in validation accuracies within 1 percent. Results well above this threshold were obtained with histogram equalization and matching, indicating that these variations were due to the image enhancements. In general, the validation accuracies during training levelled off by 40,000 iterations, indicating that the networks were trained for a sufficient number of epochs. The average difference between the validation and testing accuracy for F1-Score and IOU for BANDON for all six networks (regular, equalized, and matched with and without semantic information) was less than 1, indicating that the network did not suffer from overfitting.

4.2 Enhancement Techniques

From the visual and numerical results, it is evident that while both histogram equalization and matching can drastically improve testing accuracies, there is no method to apply universally in all cases. Instead, the decision of which enhancement to apply should depend on the spectral characteristics of the dataset.

The highest accuracies occurred when the same enhancement was used for training and testing. This result was expected since networks do not perform as well on data that differs from their training data.

In the case of S2Looking, histogram equalization was found to be particularly effective, as almost all the images were very dark before equalization. This leads to the recommendation that histogram equalization be applied to darker datasets. With satellite imagery, images tend to be darker since they are farther away from the ground. Off-nadir images also tend to be darker due to their reflectance angle. Therefore, equalization could prove invaluable in improving damage detection from off-nadir satellite imagery in a disaster scenario, as it did for the S2Looking dataset. Histogram matching decreased the accuracies for S2Looking, potentially due to some after-images becoming darker in the matching process, making it harder to detect features.

Histogram matching was effective at improving results for the Haiti dataset. One potential reason for the Haiti dataset seeing larger improvements than other datasets with matching was that the matching was performed for 512 × 512 images instead of the larger 2048 × 2048 image size of other datasets, leading to more specific colour matching. Another potential reason is that the before and after images had different spectral characteristics, particularly in the red band.

Applying enhancements to the BANDON and LEVIR-CD datasets made little difference and, in some cases, decreased the achieved accuracies. LEVIR-CD and BANDON primarily contain Google Earth images that have already undergone radiometric preprocessing, which could explain why equalization and matching did not improve accuracy. The two datasets that did not undergo prior radiometric enhancements, S2Looking and Haiti, both saw notable improvements in accuracy. In a disaster scenario with raw imagery sources, applying radiometric enhancements could be key to detecting more damage. While equalization and matching improved the results in some situations, these enhancements should be applied on a case-by-case basis.

4.3 Limitations

The Changer network trained on the BANDON change detection dataset was found to be insufficient in achieving high accuracy testing results for damage detection scenarios. The extent to which a dataset reflects damages seemed to lead to lower accuracies. For the Haiti dataset containing solely damages, the network did not classify all the pixels in each building as damaged and missed many damaged buildings entirely, likely leading to the lowest accuracies of the four datasets tested. The S2Looking dataset includes damages and changes and produced the second lowest accuracies. These results indicate that networks trained on change detection datasets are less accurate

https://doi.org/10.5194/isprs-archives-XLVIII-M-4-2024-33-2024

for damage detection tasks, underscoring the need for large damage-specific datasets for training.

One of the limitations identified with histogram equalization and matching is that the images need to contain somewhat similar information. Large differences between images can cause significant colour distortion during radiometric enhancements. Examples where images may not contain similar information include major seasonal or tidal differences. In these scenarios, histogram equalization or matching may perform worse due to them distorting the colours greatly to match the images. Future research could explore applying thresholds so that if the pixel values in the before image were a certain percentage of the full spectrum apart from pixels in the after image, histogram equalization and matching would not be applied. Histogram matching or equalization does introduce an additional step into the process. However, it does not add much processing time as it can be done for entire datasets at once in a matter of hours or on individual images in seconds.

In the case of the LEVIR-CD dataset, applying equalization and matching resulted in a decrease in overall accuracy. It is thought that perhaps because the LEVIR-CD dataset was the only nadir angle dataset, the already near-perfect alignment of the images meant that making the spectral characteristics of both images the same was a hindrance to the model's ability to detect change instead of an aid. This, coupled with the very limited improvements on the BANDON test data, which already had very high existing accuracies, suggests that while aligning the spectral characteristics of datasets is undoubtedly a useful tool to improve accuracies for datasets, its effectiveness does not lead to consistent improvements across all datasets.

4.4 Recommendations and Future Research

Subsequent research could build upon this project in many ways. The Haiti dataset was large enough for testing but not for training a network. This research outlines the need for large natural disaster-specific off-nadir datasets for training. The addition of the semantic information may not be necessary for a new dataset as it did not lead to improvements in this research. Applying radiometric enhancements to datasets and images on a case-by-case basis can lead to higher accuracies. The radiometric enhancement techniques could be tested on additional network architectures and training datasets. Thresholds and additional types of radiometric enhancements could be implemented to optimize when to use each radiometric enhancement technique depending on the spectral characteristics of the images. The loss function could be weighted to account for the class imbalance in datasets. The number of iterations for training could be increased while balancing the potential overfitting problems that are common with additional training. For areas prone to natural disasters having up-to-date imagery of the area, trained networks ready, and information about what post-disaster images may look like, could lead to selecting an appropriate radiometric enhancement technique and improved disaster management.

5. Conclusion

In this study, two radiometric enhancement techniques, histogram matching and equalization, were employed to four datasets to test whether it would have an impact on detecting changes and damages for a leading network called Changer. Histogram matching and equalization applied on a case-by-case basis to training and testing data drastically improved accuracies for two of the four datasets tested. The darker S2Looking dataset F1-Score increased by 17 when applying histogram equalization. Although the Haiti dataset had low accuracies overall, the F1-Score improved by over 20 with histogram matching. The potential of histogram matching and equalization is clear.

Radiometric enhancements were frequently applied to images before the use of convolutional neural networks became popular. In this research, applying radiometric enhancement on off-nadir imagery led to considerable accuracy improvements for a change and damage detection neural network, indicating that radiometric enhancement techniques should not be abandoned despite the advent of advanced networks. Ideally, radiometric enhancement techniques would be applied on a case-by-case basis to improve accuracy, depending on the spectral characteristics of the datasets. A leading network trained in change detection showed poor results on earthquake data, indicating the need for extensive damage-specific datasets for training. Future research could expand this concept by optimizing when to use various radiometric enhancement techniques. Continuing to pursue radiometric enhancement techniques and building damage-specific datasets could prove pivotal to improve accuracies to ultimately assist rescue, relief, and reconstruction efforts.

References

Abdi, G., Esfandiari, M., Jabari, S., 2021: Building damage detection in post-event high-resolution imagery using deep transfer learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 531–534. doi.org/10.1109/ IGARSS47720.2021.9553720

Abdi, G., Esfandiari, M., Jbari, S., 2022: A deep transfer learning-based damage assessment on post-event very high-resolution orthophotos. *Geomatica*, 75(4), 237-250. https://doi.org/10.1139/geomat-2021-0014

Chen, H., Shi, Z., 2020: A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sensing*, 12(10), 1662. doi.org/10.3390/ rs12101662.

Codegoni, A., Lombardi, G., Ferrari, A., 2023: TINYCD: a (not so) deep learning model for change detection. *Neural Computing & Applications*, 35, 8471–8486. doi.org/10.1007/s00521-022-08122-3

Fang, S., Li, K., Li, Z., 2022: Changer: Feature interaction is what you need for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 16, 1-11. doi.org/10.1109/TGRS.2023. 3277496.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M, 2019: xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. doi.org/10.48550/arXiv.1911. 09296

Holail, S., Saleh, T., Xiao, X., and Li, D., 2023: AFDE-Net: Building Change Detection Using Attention-Based Feature Differential Enhancement for Satellite Imagery. *IEEE Geoscience and Remote Sensing Letters*, 20, 1-5, Article 6006405. doi.org/10.1109/LGRS.2023.3283505.

Jabari, S., Zhang, Y., 2016: RPC-based coregistration of VHR imagery for urban change detection. *Photogrammetric Engineering & Remote Sensing*, 82(7), 521-534. doi.org/10.1016/S0099-1112(16)82059-5.

Jabari, S., Zhang, Y., 2017: Building Change Detection Improvement Using Topographic Correction Models. *Advances in Remote Sensing*, 6, 1-22. doi.org/10.4236/ars.2017.61001.

Ji, S., Wei, S., Lu, M., 2019: Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586. doi.org/10.1109/TGRS. 2018.2858817

Li, X., Ding, J., 2023: Spectral–Temporal Transformer for Hyperspectral Image Change Detection. *Remote Sensing*, 15(14), Article 14. doi.org/10.3390/rs15143561

Liu, J., Gong, M., Qin, K., Zhang, P., 2018: A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3), 545–559. doi.org/10.1109/TNNLS.2016.2636227

Liu, T., Yang, L., Lunga, D., 2021: Change detection using deep learning approach with object-based image analysis. *Remote Sensing of Environment*, 256, 112308. doi.org/10.1016/j.rse.2021.112308

Meta AI, 2023: LEVIR-CD Benchmark (Building change detection for remote sensing images). paperswithcode.com/sota/ building-change-detection-for-remote-sensing

Pang, C., Wu, J., Ding, J., Song, C., Xia, G. S., 2023: Detecting building changes with off-nadir aerial images. *Science China Information Sciences*, 66(4), 140306. doi.org/10.1007/s11432-022-3691-4.

Shen, L., Lu, Y., Chen, H., Wei, H., Xie, D., Yue, J., Chen, R., Lv, S., & Jiang, B, 2021. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sensing*, 13, 5094. doi.org/10.3390/rs13245094.

Wang, J., Meng, L., Li, W., Yang, W., Yu, L., & Xia, G.-S., 2022: Learning to Extract Building Footprints From Off-Nadir Aerial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 1294–1301. https://doi.org/ 10.1109/TPAMI.2022.3162583

Zhang, J., Shao, Z., Ding, Q., Huang, X., Wang, Y., Zhou, X., Li, D., 2023: AERNet: An Attention-Guided Edge Refinement Network and a Dataset for Remote Sensing Building Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 1-16, 2023, Article 5617116. doi.org/10.1109/TGRS.2023.3300533.