Journal of Theoretical and Applied Information Technology
15th January 2025. Vol.103. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL DEEP LEARNING METHOD FOR DETECTING CHANGES IN SATELLITE IMAGERY

[1]DIAA HAFEZ IBRAHIM, [2] REDA A. EL-KHORIBI, [3] FARID ALI MOUSA

[1] Electrical and Electronics Communication Department, Faculty of Engineering, October University for Modern Sciences and Arts (MSA University), Giza, Egypt
[2] Information Technology Department, Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt.
[3] Information Technology Department, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, Egypt

Email: [1] dhafez@msa.edu.eg, [2] ralkhoribi1@gmail.com, [3] farid.cs@gmail.com

ABSTRACT

The identification of change information has been crucial to the use of satellite imagery, the monitoring of land cover and use, the estimation of damage from natural catastrophes and the detection of military targets. There are numerous conventional techniques for detecting changes in multispectral remote sensing images, but they frequently fall short of our needs for durability, accuracy, and precision. This paper introduces a novel deep learning method for identifying changes in satellite data, with an emphasis on environmental changes and urban expansion. Using 24 pairs of Sentinel-2 satellite image data collected from 2015 to 2018, of which 10 pairs were used for testing and 14 pairs for training. Thirteen spectral bands with different spatial resolutions (10 m, 20 m, and 60 m) make up each multispectral image pair. The paper evaluates changes in urban and rural environments using the visible spectrum bands (2, 3, and 4) at a resolution of 10 m. The collection contains manually annotated changes.

The paper compares the results obtained from the proposed solutions by Siamese Network and U-Net to address this change detection problem. With an accuracy of 0.86, the Siamese Network is used to detect high-level structural changes between pre- and post-event images by learning similarities between paired images. With an accuracy of 0.84, U-Net, which is intended for semantic segmentation, offers pixel-level predictions that improve change detection detail while the hybrid method for pixel-level change detection that combines the Siamese Network and U-Net in order to increase accuracy even further. This approach, which uses the Siamese Network for patch-wise similarity comparison and U-Net for fine-grained pixel segmentation, yields the maximum accuracy of 0.91. An efficient framework for applications in urban planning, crisis management, and environmental monitoring is suggested by the suggested hybrid technique, which shows great promise for accurate and thorough change detection in satellite imagery

Keywords *Satellite Images, U-net, Siamese, Urban development, Change detection*

1. INTRODUCTION

One of the classic problems in satellite imaging is change detection. Cloud detection, road and building appearance, deforestation monitoring, and agricultural crop monitoring are only a few examples of the changes that need to be examined. [1]

Numerous applications make considerable use of change detection techniques, such as urban change analysis [2], environmental monitoring [3], land management [4], and disaster assessment [5]. The frequency of major climate change-related disasters, including heat waves, storms, floods, and drought, has simultaneously exposed a new research issue and the need for more efficient automated change detection techniques.

Deep learning has been introduced for change detection in remote sensing and has demonstrated good performance, driven by the afore mentioned observations.

Numerous reviews that concentrate on deep learning for data from remote sensing have recently been released. The deep learning methods used in all of the main sub-areas of remote sensing, such as classification, restoration, denoising, target recognition, scene understanding, and other tasks, have been compiled in this research.

Satellite image processing poses a number of unique issues, mostly because of the images' unusual features in comparison to traditional images.[6]

First, a major constraint is frequently the size of the items that are detected. A football stadium, for instance, may take up hundreds of pixels in a traditional photograph, enabling in-depth examination. A satellite image with a resolution of 10 meters per pixel, on the other hand, would only use a small number of pixels to depict the same stadium, reducing spatial information and making object detection and categorization more difficult.

Second, the spectral composition of satellite photographs is very different from that of conventional images. Satellite photos can have hundreds of spectral bands, but conventional photographs usually have three color channels (red, green, and blue). This high dimensionality makes data processing more difficult and calls for sophisticated feature extraction and analysis methods.

Third, further unpredictability is introduced by the sun's angle at the moment of image acquisition. When comparing two photographs, shadows cast by large objects like buildings and mountains can show up as abnormalities or alterations, especially when utilizing simple image processing methods. This fluctuation complicates the detection of true changes over time.

Lastly, satellite imagery is greatly influenced by the ambient circumstances at the moment of acquisition. Features in the image may be obscured by elements like snow, fog, and clouds; therefore, advanced preprocessing techniques are necessary to counteract these effects and retrieve relevant information.

These difficulties demonstrate the necessity of sophisticated algorithms and reliable processing and analysis techniques for satellite images, especially for applications like environmental monitoring, land-use classification, and disaster management.

A fundamental technique in earth observation, this procedure aims to differentiate between altered and unaltered pixels in bi-temporal or multi-temporal remote sensing images taken from the same geographic region or area but at various periods, respectively [7]. The primary goal of the change detection system is to assign a binary label to each pixel based on a pair or sequence of co-registered images. Accordingly, a null label indicates an area that has not changed, but a positive label indicates that the area of that pixel has changed. In reality, change detection is an effective technique for multi-temporal analysis, urban mapping, and video monitoring.

Numerous applications have effectively employed change detection. Change detection is specifically used in the agricultural industry for disaster assessment, shifting cultivation monitoring, and deforestation monitoring. It is currently used in the military to gather data on battlefield location, enemy military force movements, new military facilities, and damage assessments [8]. Change detection is employed in the civil field to regulate the growth of urban areas and the expansion of cities [8]. Additionally, it is used to track the effects of climate change, which are typically linked to rising greenhouse gas (GHG) emissions in the atmosphere. Examples of these effects include changes in sea level and glacier facies and mass balance.

Traditional Techniques for Change Detection in Satellite Imagery

Several fundamental methods have been developed to detect changes in two satellite images captured at different times. These methods primarily focus on analyzing pixel-level variations between spatially registered images and have been widely used in remote sensing applications.[9]

1. Image Differencing

Image differencing involves subtracting the pixel values of two spatially registered images from different acquisition dates. The resulting difference image highlights regions where significant variations occur, indicating areas of change. One of the advantages of this method is its simplicity and speed, making it an efficient way to quickly identify areas that have undergone change. However, the primary disadvantage is that it is highly sensitive to noise and can struggle to differentiate between actual changes and minor variations caused by atmospheric conditions or sensor discrepancies. Equation 1 describe image difference

$$ID = |Image_i^2 - Image_i^1| \quad (1)$$

Such that $Image_i^2$ is image 2 / Number of bands and $Image_i^1$ is image 1 / Number of bands

2. Image Rationing

Image rationing calculates the ratio of pixel values from two registered images on a band-by-band basis. In areas where change has occurred, the ratio values will significantly deviate from 1, either being higher or lower depending on the type of change. This method as show in equation 2 is advantageous because it offers a simple and rapid way to detect changes. However, it has been criticized for producing non-normal histogram distributions, which can limit its effectiveness in detecting subtle changes and make interpretation more challenging, especially in areas with low contrast or gradual transitions.

$$IR = \left|\frac{Image_i^2}{Image_i^1}\right| \quad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. Normalized Difference Built-Up Index

The NDBI technique involves calculating a Built-Up Index for each image date using the Near Infrared (NIR) and Shortwave Infrared (SWIR) bands. Equation 3 describes the mathematical foundation of NDBI.

$$NDBI = \frac{SWIR-NIR}{SWIR+NIR} \quad (3)$$

The difference between these indices across the two dates helps highlight changes in urban environments, particularly the expansion of built-up areas. One of the major advantages of this method is that it emphasizes the spectral differences associated with urban features, reducing the influence of topographic effects and varying illumination. However, it can also enhance random or coherence noise, which can complicate the interpretation of changes and lead to false positives in non-urban areas.

4. Average Intensity

The AI method normalizes the images by calculating the average intensity of pixel values between two images as shown in equation 4.

$$AI = \frac{Image_i^2 - Image_i^1}{Image_i^2 + Image_i^1} \quad (4)$$

It is particularly effective in identifying changes in shaded areas where the average brightness is lower. The advantage of this method lies in its ability to detect changes in regions with low average brightness, such as areas under shadow or in dense vegetation. However, when environmental conditions vary significantly, this method requires additional operators, often based on statistical models, to adapt and produce accurate results, making it more complex to implement under varying conditions.

5. Euclidean Distance

The Euclidean Distance method measures the change by calculating the spatial distance between pixel values in a multi-dimensional spectral space as shown in equation 5.

$$ED = \sqrt{\sum_{i=1}^{n} (Image_2^i - Image_1^i)^2} \quad (5)$$

Larger distances indicate a higher likelihood of change, while smaller distances suggest no change. The advantage of this method is its ability to quantify changes in a way that is straightforward to interpret. However, the main disadvantage is that it may not perform well when there are subtle changes or when spectral overlaps exist between different features in the images, leading to ambiguity in the change detection results.

6. Image Regression

Image regression establishes relationships between two images by fitting a regression model to the pixel values of each corresponding pair of images. The second image's pixel values are then estimated using the regression function, and the difference between the regressed values and the actual pixel values is used to detect changes as shown in equation 6.

$$IR = a\,I + b \quad (6)$$

Such that I is the first image, a is the slope and b in intercept.

One of the key advantages of this method is that it helps reduce the impact of atmospheric, sensor, and environmental differences between the two images, leading to more reliable change detection. However, a major limitation is that it requires the development of accurate regression functions for the selected spectral bands, which can be time-consuming and complex, particularly when the images come from different sensors or have different radiometric characteristics.

These traditional techniques have their strengths and weaknesses, often facing challenges related to noise sensitivity, environmental variability, and sensor differences. While they offer straightforward and computationally efficient methods for detecting changes, their limitations have driven the development of more advanced machine learning and deep learning models, which are capable of overcoming many of these challenges by leveraging more sophisticated feature extraction and analysis techniques.[9]

The structure of this publication is as follows: Section 2 summarized earlier research. Section 3 describes the proposed method. Section 4 discusses the results and their implications. The conclusion is ultimately discussed in section 5.

2. PREVIOUS WORK

Pixel-based and object-based techniques are the two categories of traditional change detection techniques. Studies mostly concentrate on the analysis of difference images (di) using pixel-based approaches find in [10] by Kelly,. et al.

Object correlation analysis is used in image segmentation by Yuan et al. [11]. Information about the segmented objects is also used to detect changes. Wang et al. [12] segment multitemporal images using a modified seed-region growing technique. Geostatistical characteristics are extracted via multiresolution segmentation in [12]. These approaches require distinct processes including feature extraction and classification, and segmentation is one of several steps.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Deng et al. [13] propose an architecture that combines a Mask R-CNN network for object-based instance segmentation with a multi-scale fully convolutional network for pixel-based semantic segmentation. This hybrid approach leverages the strengths of both networks to improve the accuracy of both instance-level and semantic-level change detection tasks.

The field of remote sensing image processing has become interested in deep learning in recent years. Numerous studies on change detection based on deep learning have been conducted. For change detection, Arosio et al. [14] use a fully linked deep neural network. One pixel's neighbourhood serves as input, and the pixel's classification outcome serves as output. A bipartite differential neural network is proposed by Zhao et al. [15], wherein the input image pair is superimposed with two change disguise maps (cdm) that identify the altered regions. The network output is used to build an objective function that is optimized. Small-scale datasets are used to train these models.

However, our model uses a Siamese architecture, where the two pictures are passed through in turn, with the bottom levels sharing weights. The procedure of learning similar features through shared weights makes sense because the two photos were taken in the same scene at separate times. A Siamese convolutional network for change detection is suggested in [16]. The model of [16] uses a straightforward thresholding segmentation based on Euclidean distance that is distinct from the network while combining the data acquired by the Siamese CNN. We create deeper modules in our model to improve information fusion and segmentation. Due to its powerful ability to leverage contextual information, the fully convolutional network [17] is the most widely used convolutional model for semantic image segmentation.

Change detection tasks involve generating a change map from a pair of input images and assigning pixel-wise semantic change labels. In recent years, numerous deep learning-based methods have been proposed for change detection in satellite imagery. These methods are primarily derived from classification or semantic segmentation techniques. Daudt et al. [18] were among the first to introduce an end-to-end framework for change detection using Convolutional Neural Networks (CNNs), which predicted the label of the central pixel in each image patch. However, this approach overlooked the context surrounding the central pixel, which is crucial for detecting small objects. In a subsequent study, Daudt et al. [18] proposed a fully convolutional encoder-decoder network, where the encoder utilized a Siamese network, and for the first time, skip connections were integrated within the Siamese network to improve feature sharing and context preservation.

An encoder and a decoder make up an encoder-decoder model. The decoder retrieves the details and produces the segmentation map after the encoder records high-level information along the spatial dimension of feature maps. Deconvolution is used as the high-level feature recovery in. [19]. Using skip connections, U-Net [20] connects the encoder features to the same-level decoder output.

This part offers a thorough analysis of both conventional and contemporary change detection methodologies, clearly showing the transition from object-based and pixel-based approaches to deep learning-driven strategies. While the inclusion of hybrid approaches, such as the integration of Mask R-CNN and fully convolutional networks, demonstrates the potential of combining methodologies for improved performance, the thorough discussion of traditional methods, such as segmentation and feature extraction, highlights their complexity and limitations. Additionally, the examination of deep learning methods, specifically the use of encoder-decoder frameworks like U-Net and Siamese architectures, shows a notable advancement in the use of contextual and spatial information for increased accuracy. The insights in this section could be strengthened, nevertheless, by a more thorough evaluation of these methods' computational effectiveness, scalability, and performance on other datasets.

3. DATA SET AND PROPOSED MODEL

3.1 Dataset

The Open and Standardized Change Detection (OSCD) dataset was created to provide a trustworthy standard by which to evaluate different change detection techniques. The dataset, which was created with an emphasis on urban regions, solely designates structural changes and urban growth as "Change," ignoring natural oscillations like tide shifts or vegetation growth. Researchers and practitioners interested in change detection can use this dataset, which supports the development and assessment of single-band, color, and multispectral change detection algorithms by providing pixel-level change labels for supervised learning models.

Sentinel-2 satellite photos, which include wavelengths from ultraviolet to short-wave infrared, and 13 spectral bands at resolutions ranging from 10 m to 60 m, are used in the OSCD collection. To include locations with discernible urbanization shifts, twenty-four regions from around the world—each measuring roughly 600 by 600 pixels at a resolution of 10 m—were chosen. The Medusa toolkit is used to trim each image region's 26 images (13 bands for each pair) to precise coordinates.[21]

Challenges and Limitations

The OSCD dataset has a number of significant drawbacks, despite being a useful tool for applying supervised learning techniques and methodically evaluating change detection systems. First, the Sentinel-2 satellite can record important structural changes, such as the appearance of massive buildings, even though its photos have a relatively poor quality. Finer alterations, including adding lanes to roads, extending old structures slightly, or adding new buildings, are frequently harder to spot. As a result, analysts' subjective interpretations of these minute changes may cause differences in manually created change maps.

Using OpenStreetMap data from various dates was one method that was used to automate the creation of change maps. This technique sought to identify changes in an organized, automated manner by comparing map data that corresponded to the dates of each image pair. However, a number of drawbacks made it ineffectual. First off, a lot of map modifications indicated additions that weren't necessary constructed between the dates of image acquisition, which led to erroneous temporal alignment. Furthermore, the accuracy of dating earlier OpenStreetMap data was limited because, until 2017, maps frequently only offered one version annually, making accurate temporal matching difficult.

Additionally, the collection is limited to photos taken starting in June 2015 because the Sentinel-2 spacecraft was launched in 2015. This restricts the range of discernible changes by limiting the dataset's temporal scope to roughly two and a half years or fewer for many image pairs. Furthermore, there are more "no change" pixels in the dataset than "change" pixels, which results in an unbalanced label distribution that could impair algorithm performance. When using the OSCD dataset for change detection investigations, several aspects should be taken into account.

3.2 Proposed Model

Change detection in satellite imagery is critical for understanding urban dynamics, monitoring natural disasters, and managing resources effectively. the paper explores and evaluates three distinct methodologies—Siamese network, u-net, and a hybrid approach combining both—for identifying structural changes in urban environments using the open and standardized change detection (OSCD) dataset. these methods leverage the rich spectral and spatial information provided by sentinel-2 multispectral satellite imagery to achieve pixel-level accuracy in change detection tasks.

The following sections detail the experimental workflows for each methodology, emphasizing the preprocessing techniques, network architectures, and evaluation strategies employed. By comparing these approaches, this study aims to identify the most effective method for urban change detection in high-resolution satellite imagery.

3.2.1 Change Detection Using a Siamese Network

A multi-step process of data pre-processing, network architecture design, and feature comparison is used in the methodology for applying a Siamese Network to change detection in the OSCD (Open and Standardized Change Detection) dataset. This approach is optimized for identifying changes in urban structure. Sentinel-2 multispectral satellite images covering wavelengths from the visible to short-wave infrared spectrum are included in the OSCD dataset. The bands were taken at different spatial resolutions (10 m, 20 m, and 60 m). All photos are resampled to a standard 10 m resolution in order to guarantee spatial consistency between these bands, enabling consistent pixel alignment across channels. Pixel-by-pixel comparisons inside the network are made possible by this pre-processing phase, which also guarantees compatibility across spectral bands. Critical spectral information for urban characteristics is then captured by structuring the resampled images into multi-channel inputs using certain band combinations, such as Red, Green, Blue (RGB), and Near-Infrared (NIR). Given the dataset's emphasis on urban evolution, this band selection prioritizes structural components while reducing vulnerability to natural variability.

Journal of Theoretical and Applied Information Technology
15th January 2025. Vol.103. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

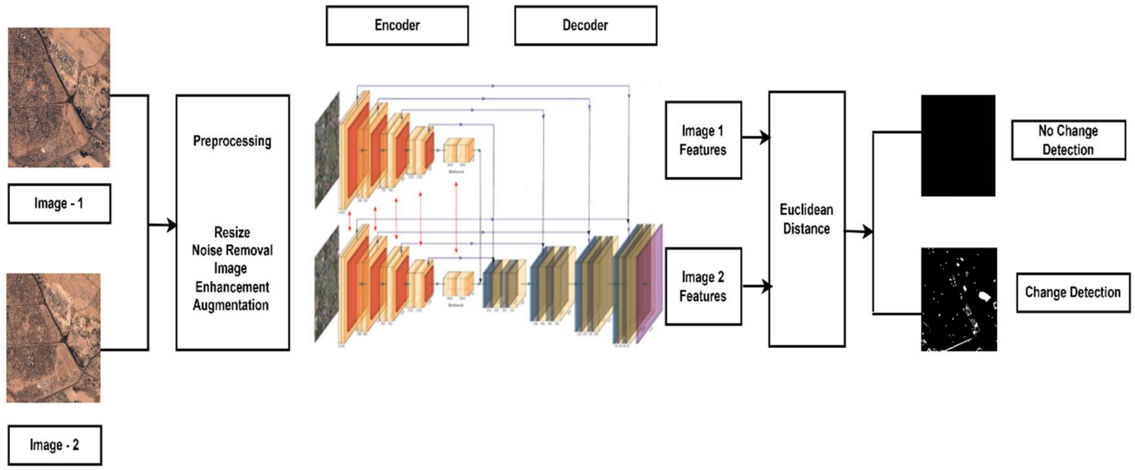

Figure 1. Change detection using Siamese Network

As shown in Figure 1. The application of a Siamese Network for change detection involves a systematic process divided into three key phases: **data preprocessing**, **network encoding and decoding**, and **feature vector comparison for change detection**. This methodology ensures accurate identification of structural changes between pre-event and post-event satellite images.

In the **data preprocessing phase**, paired pre-event and post-event images undergo several preparatory steps to enhance their quality and maintain consistency. First, all images are resized to a uniform resolution to ensure spatial alignment between the paired inputs. This alignment is crucial for pixel-to-pixel comparisons during the subsequent stages. Next, noise removal techniques are applied to mitigate distortions caused by environmental factors such as clouds, shadows, or atmospheric interference, thereby enhancing the clarity of structural details. Additionally, image enhancement methods, such as contrast stretching and histogram equalization, improve the visibility of key features critical for detecting urban changes. To further enhance the robustness of the model, data augmentation techniques, including rotations, flips, and scaling, are applied. These augmentations expand the diversity of the dataset and improve the model's ability to generalize to unseen data, reducing overfitting risks.

The **network encoding and decoding phase** is the core of the Siamese Network architecture. The network consists of two identical sub-networks (encoders) with shared weights that independently process the pre-event and post-event input images. Each sub-network, typically a convolutional neural network (CNN), extracts high-level feature representations of the input patches, capturing both spatial and spectral characteristics. The shared weights ensure that both images are processed identically, enabling meaningful feature comparisons. Following the encoding process, a decoding stage refines the extracted feature vectors, reconstructing contextual information while preserving spatial relationships, which is essential for precise analysis.

In the final **feature vector comparison phase**, the pre-event and post-event feature vectors generated by the Siamese Network are compared using a distance metric, such as the Euclidean distance. This metric quantifies the dissimilarity between the feature vectors, reflecting the extent of changes between the two images. A high Euclidean distance indicates significant differences between the feature representations, suggesting the presence of structural changes, while a low distance implies minimal or no changes. Based on the similarity scores, a binary classification is performed for each image patch, identifying it as either "changed" or "unchanged." The results are aggregated to produce a change map that visually highlights the regions where changes have occurred.

This structured approach leverages the computational efficiency and feature comparison capabilities of the Siamese Network, making it well-suited for change detection tasks in urban environments. By focusing on feature-level differences, the methodology enables precise identification of structural changes while minimizing processing demands.

The Siamese Network uses the Euclidian distance metric to compare feature vectors from the pre- and post-event patches. By quantifying the degree of change in feature space, this distance metric—which is the absolute difference between

Journal of Theoretical and Applied Information Technology
15th January 2025. Vol.103. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the feature vectors—allows the network to gauge how much has changed between patches. The network can distinguish between stable and changed regions by using the Euclidian distance to identify both significant and subtle structural changes. After passing the resultant distance metric to a fully connected layer, a softmax classifier produces a binary classification of "Change" or "No Change." This layer directly addresses the goal of detecting structural changes by allowing the network to link variations in feature vectors to a change detection decision.

Binary cross-entropy loss, a loss function that works well for binary classification tasks—where the goal is to reduce the error between the true and predicted labels—is used to train the model. The network's prediction is specifically compared against ground truth labels, which are given as binary indications of change in the OSCD dataset, using the cross-entropy loss. The Siamese Network may learn discriminative characteristics that differentiate natural variations, like seasonal changes in flora, from urban alterations by training on labeled patch pairings. While maintaining insensitivity to incidental environmental variables, this training procedure guarantees that the network gains a solid comprehension of structural changes.

The Siamese Network creates change maps for fresh pre- and post-event picture pairs following training. The resulting change maps are subjected to Gaussian smoothing techniques and morphological filters (such as closing and opening operations) in order to improve these predictions and lower noise. By lowering artifacts and improving interpretability for additional analysis, these post-processing procedures raise the change maps' visual quality and accuracy.

Using a Siamese Network to efficiently record and compare structural changes in high-resolution satellite pictures, this methodology offers a thorough approach to urban change detection. The model is a useful tool for urban monitoring and analysis jobs since it produces dependable findings through meticulous data pre-processing, feature extraction, and post-processing.

3.2.2 Change Detection Using U-Net

In order to accurately detect structural changes in urban areas using high-resolution satellite imagery, the methodology for implementing U-Net for change detection on the OSCD dataset consists of a meticulously planned workflow that includes data pre-processing, U-Net architecture setup, and post-processing. Multispectral Sentinel-2 photos with resolutions ranging from 10 m to 60 m over 13 spectral bands, covering a range of wavelengths from visible to short-wave infrared, make up the OSCD collection. All photos are resampled to a standard resolution of 10 m in order to preserve spatial alignment and uniformity across bands. In order to enable precise pixel-level comparison during segmentation, these resampling guarantees consistency across spectral channels.

As shown in Figure 2. The implementation of U-Net for change detection follows a structured workflow divided into three main phases: **data preprocessing**, **localization using U-Net**, and **change detection**. This approach leverages the segmentation capabilities of U-Net to achieve accurate pixel-level identification of structural changes in satellite imagery.

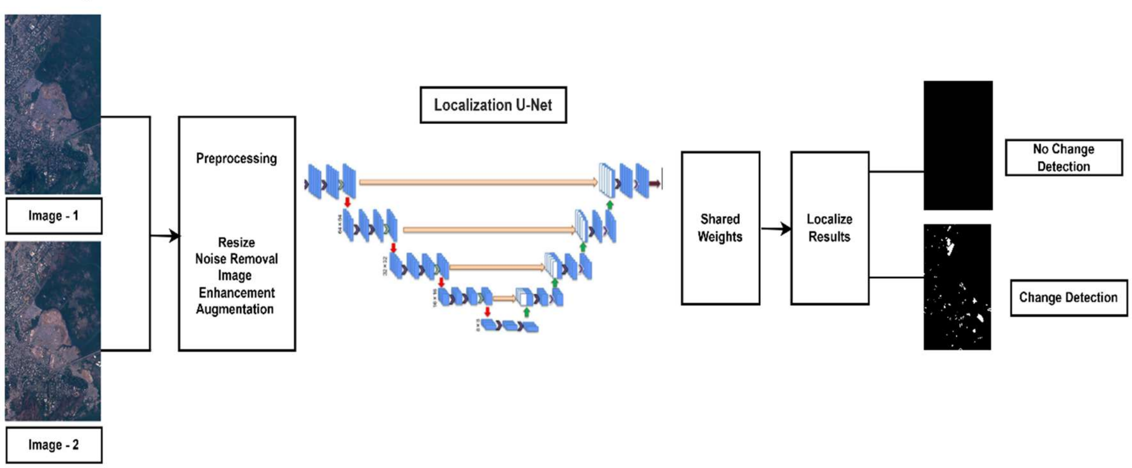

Figure 2. Change detection using U-Net

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In the **data preprocessing phase**, the input images undergo several preparatory steps to ensure uniformity and enhance data quality. All images are resized to a standard resolution to maintain spatial consistency across the dataset, facilitating pixel-level alignment. Noise removal techniques are applied to address potential distortions caused by environmental conditions such as clouds, shadows, or atmospheric interference, ensuring the visibility of key structural details. Additionally, image enhancement methods, such as contrast adjustment and histogram equalization, improve the clarity and contrast of the images, making features more discernible. Data augmentation techniques, including random rotations, flips, and scaling, are employed to increase the diversity of the training dataset, enhancing the model's generalizability and robustness against overfitting.

The **localization phase** utilizes the U-Net architecture to generate a preliminary segmentation map that highlights potential regions of change. The U-Net model, based on an encoder-decoder structure, localizes changes by capturing both high-level contextual information and fine-grained spatial details. During the training process, U-Net generates standard weights optimized for the specific dataset, enabling consistent and reliable feature extraction. These weights are subsequently used to process the input data, localizing changes by segmenting areas that exhibit notable structural variations. The model's ability to focus on relevant features ensures accurate delineation of regions of interest.

In the **detection phase**, the results from the U-Net localization are analyzed to confirm and refine the identified changes. The segmented output is post-processed to eliminate false positives and noise, ensuring that only genuine changes are retained. This step includes thresholding techniques and morphological operations to refine the boundaries of detected regions. The final output is a detailed change map that highlights structural differences between pre-event and post-event images at a pixel level.

This U-Net-based methodology combines the advantages of robust preprocessing, precise localization, and refined detection to deliver highly accurate change detection results. Its ability to capture spatial and contextual features makes it particularly effective for identifying subtle structural changes in urban and natural environments.

Because the U-Net architecture combines an expansive path for accurate localization and a contracted path for feature extraction, it is especially well-suited for pixel-level segmentation. The input patch is downsampled via a sequence of convolutional and max-pooling layers in the contracting route, each of which captures progressively abstract information. The compressed feature maps are then upsampled using the expansive path, which uses skip connections from matching layers in the contracting path to recover fine-grained features and spatial resolution. Because it allows U-Net to preserve contextual information while generating high-resolution change maps, this symmetrical design is crucial for change detection.

For training, the model takes two multi-channel patches—representing the same region from pre-event and post-event times—and stacks them as input. This input format allows the network to analyze temporal differences within each patch and learn features indicative of structural changes. During the upsampling phase, U-Net merges these temporally derived features, producing a pixel-wise probability map where each pixel represents the likelihood of change or no change. The output layer is a softmax or sigmoid activation, depending on the configuration, to generate binary labels (change/no change) for each pixel, enabling the model to produce highly localized change maps.

The model is optimized using binary cross-entropy loss, as this loss function is well-suited for binary pixel-wise segmentation tasks. During training, each output pixel is compared with a labeled ground truth from the OSCD dataset, with the loss function minimizing discrepancies between predicted and actual change labels. Training on labeled patch pairs from pre- and post-event times allows U-Net to develop a nuanced understanding of change patterns, capturing urban growth and structural transformations while ignoring natural variations like vegetation changes.

This approach is quite successful in detecting structural changes in urban areas because it makes use of U-Net's segmentation strengths for accurate, pixel-level change identification. This method produces precise and comprehensive change maps by combining meticulously planned pre-processing, U-Net architecture, and post-processing, offering important insights for jobs involving urban monitoring and spatial analysis.

3.2.3 Hybrid Approach

In order to improve accuracy, the hybrid method that combines a Siamese Network with U-Net for pixel-level change detection makes use of both fine-grained pixel-level segmentation and patch-wise similarity comparison. By using the Siamese Network to pinpoint areas where changes are most likely to have taken place, this technique lowers processing demands by letting the U-Net

Journal of Theoretical and Applied Information Technology
15th January 2025. Vol.103. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

concentrate just on these potential areas for in-depth segmentation. This method is computationally efficient and offers more accurate change detection by integrating the advantages of both systems.

To guarantee consistency and relevancy of inputs, pre-processing procedures are carried out on the Sentinel-2 photos in the first step of this workflow. In order to facilitate pixel-to-pixel alignment between spectral channels, all images are resampled to a single resolution because Sentinel-2 records multispectral data at several resolutions (10 m, 20 m, and 60 m). Each satellite image is resampled and then split into tiny patches, usually measuring 64 by 64 pixels, which serve as the Siamese Network's input. Red, Green, Blue (RGB), and Near-Infrared (NIR) are examples of relevant spectral bands that are chosen as input channels because they offer crucial information for identifying structural changes in metropolitan environments while lowering sensitivity to non-structural variables like vegetation.

As shown in Figure 3. pairs of patches taken from pre-event and post-event photos are compared for patch-wise similarity using the Siamese Network stage. High-level feature representations are produced by each patch pair passing through identical sub-networks (such as CNN-based architectures with shared weights). The network calculates a similarity score for every pair of patches using a distance metric, like L1 or Euclidean distance, which represents the extent of change between the two images. A coarse change map is produced by designating patches with high similarity scores as potential change zones. This map provides a more targeted and computationally efficient input for further research by highlighting areas with possible changes and removing stable regions.

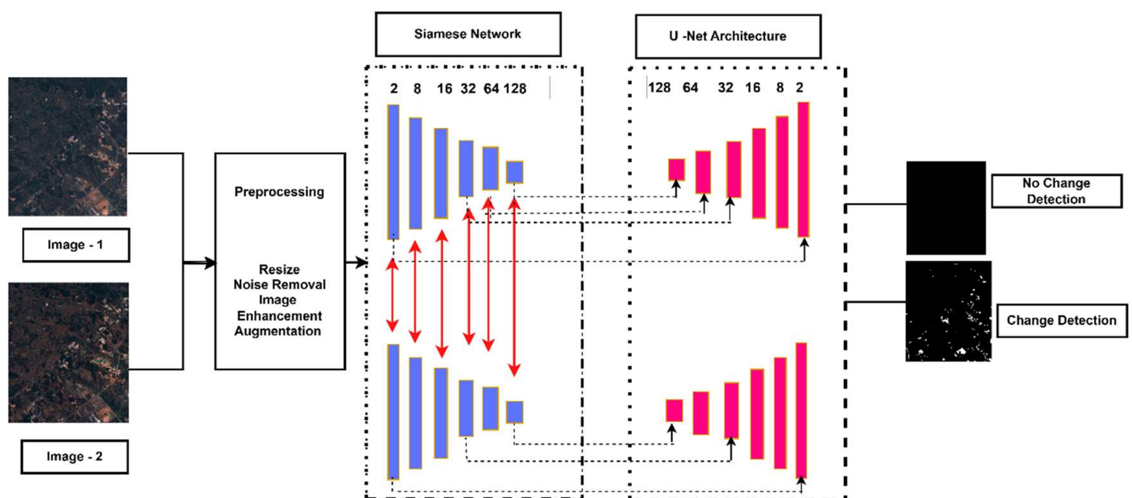

Figure 3. Proposed Method For Change Detection In Satellite Imagery

U-Net is used to fine-tune these regions at the pixel level after the initial candidate identification. By omitting sizable regions with negligible alterations, the Siamese Network's candidate regions are the only ones sent to the U-Net, lowering the computational load. U-Net learns the precise variations between pixels over time by receiving a stacked input of pre-event and post-event patches for each candidate region. With its encoder-decoder structure, the U-Net architecture carries out thorough segmentation, generating a binary change map that designates every pixel as either "Change" or "No Change." This approach allows U-Net to identify small changes with pixel-level precision by preserving spatial information and capturing context through skip links between the encoder and decoder layers.

Post-processing is the last stage of the workflow, which improves the output change map's quality and usefulness. By filling in the gaps and eliminating tiny noisy areas, morphological filtering techniques like erosion and dilation lower the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

frequency of false positives and enhance continuity in the segmented sections. A smooth view of the identified changes throughout the entire image is then provided by the reconstruction of the complete change map using the segmented patches.

The hybrid approach offers several advantages. By narrowing down candidate regions, the Siamese Network minimizes the area that U-Net needs to process, reducing computational load. The U-Net, in turn, provides high-resolution segmentation, allowing the method to handle complex changes and detect small, localized alterations that may not be apparent with patch-wise comparison alone. Post-processing filters also make this approach robust to noise, resulting in more reliable and accurate change maps.

The architecture of this hybrid method leverages two key components: the Siamese Network, with twin convolutional sub-networks for patch-wise similarity assessment, and U-Net, an encoder-decoder model that refines candidate regions with precise segmentation. Post-processing steps further enhance the final change map by filtering out noise, thereby improving overall accuracy and reliability.

With this combination, the hybrid Siamese-U-Net model offers a well-rounded solution for change detection in high-resolution satellite imagery, attaining great accuracy and computational efficiency.

4. EXPERIMENTAL RESULTS

The experimental results demonstrate the performance of three methodologies—Siamese Network, U-Net, and the proposed hybrid model combining these two approaches—for satellite image change detection. The accuracy values achieved by each method highlight in Table 1. their individual strengths and the advantages of their integration in the hybrid model. Specifically, the Siamese Network achieved an accuracy of **0.86**, the U-Net recorded **0.84**, and the proposed hybrid model outperformed both with an accuracy of **0.91**.

Table 1: Experiments Results For OSCD Dataset

Method	Accuracy
Siamese Network	0.86
U-Net	0.84
Proposed Method	0.91

The **Siamese Network** proves highly effective for identifying changes by leveraging its capability to compare pre-event and post-event image patches through feature similarity. Its architecture, which uses paired inputs and shared weights, ensures robust feature extraction and direct comparison of pixel-level changes. This approach is particularly advantageous in scenarios where localized changes must be identified efficiently, as the Siamese Network minimizes computational overhead by focusing on pairwise feature comparison. However, while it excels in detecting coarse changes, its performance may be limited in capturing complex spatial details due to its focus on patch-level similarity rather than detailed segmentation.

The **U-Net**, on the other hand, demonstrates its strength in pixel-level segmentation, making it highly suitable for capturing fine-grained spatial details and subtle structural variations in satellite images. Its encoder-decoder architecture efficiently extracts high-level features and refines them into detailed segmentation maps, enabling precise delineation of change regions. Despite its ability to localize changes accurately, U-Net may encounter challenges in handling noise or ambiguous features in large-scale satellite datasets, which could slightly impact its overall accuracy compared to Siamese Network.

By integrating the strengths of both approaches, the **proposed hybrid model** achieves superior accuracy, leveraging the Siamese Network for effective localization and U-Net for detailed segmentation. In this hybrid setup, the Siamese Network first identifies regions likely to have undergone change, reducing computational demands by narrowing the focus to relevant areas. Subsequently, the U-Net processes these regions with its segmentation capabilities, capturing intricate spatial details and producing a refined change map. This synergy between the two models mitigates their

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

individual limitations, allowing the hybrid approach to achieve the highest accuracy of **0.91**.

The results underline the benefits of combining complementary deep learning architectures to enhance change detection in satellite imagery. The hybrid model's superior performance demonstrates its potential as an effective and scalable solution for monitoring urban dynamics, environmental changes, and disaster impacts.

5. CONCLUSIONS

This study demonstrates the efficacy of deep learning models—Siamese Network, U-Net, and a hybrid approach combining the two—for change detection in satellite imagery, using the OSCD dataset as a benchmark. The experimental results reveal that while the Siamese Network excels in efficiently identifying localized changes with an accuracy of 0.86 and U-Net delivers robust pixel-level segmentation achieving an accuracy of 0.84, the proposed hybrid model outperforms both with an accuracy of 0.91.

The analysis highlights the complementary strengths of the two models: the Siamese Network's ability to localize change areas through pairwise feature comparisons and U-Net's capability to capture intricate spatial details through its encoder-decoder architecture. The hybrid model effectively integrates these strengths, enabling precise change localization followed by refined segmentation, which addresses the limitations of each individual method.

The proposed hybrid approach demonstrates significant potential for enhancing urban monitoring, disaster assessment, and environmental change detection. Its ability to provide detailed and accurate change maps makes it a valuable tool for practical applications in remote sensing and geospatial analysis. Future work could extend this research by addressing dataset limitations, such as label imbalance and temporal constraints, and exploring the integration of additional spectral bands or spatial contexts to further improve model performance. This study underscores the importance of leveraging complementary deep learning architectures to advance the field of satellite image change detection.

REFERENCES

[1]. Ma, Y., Chen, S., Ermon, S., & Lobell, D. B. (2024). Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, *301*, 113924.

[2]. Stilla, U., & Xu, Y. (2023). Change detection of urban objects using 3D point clouds: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *197*, 228-255.

[3]. Yuan, S., Li, Y., Bao, F., Xu, H., Yang, Y., Yan, Q., ... & Lin, J. (2023). Marine environmental monitoring with unmanned vehicle platforms: Present applications and future prospects. *Science of The Total Environment*, *858*, 159741.

[4]. AbdelRahman, M. A. (2023). An overview of land degradation, desertification and sustainable land management using GIS and remote sensing applications. *Rendiconti Lincei. Scienze Fisiche e Naturali*, *34*(3), 767-808.

[5]. Mansour, R. F., & Alabdulkreem, E. (2023). Disaster Monitoring of Satellite Image Processing Using Progressive Image Classification. *Computer Systems Science & Engineering*, *44*(2).

[6]. Guerri, M. F., Distante, C., Spagnolo, P., Bougourzi, F., & Taleb-Ahmed, A. (2024). Deep learning techniques for hyperspectral image analysis in agriculture: A review. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 100062.

[7]. Dimitrovski, I., Kitanovski, I., Kocev, D., & Simidjievski, N. (2023). Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *197*, 18-35.

[8]. Zhu, Q., Guo, X., Li, Z., & Li, D. (2024). A review of multi-class change detection for satellite remote sensing imagery. *Geo-spatial Information Science*, *27*(1), 1-15.

[9]. Cheng, G., Huang, Y., Li, X., Lyu, S., Xu, Z., Zhao, H., ... & Xiang, S. (2024). Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sensing*, *16*(13), 2355.

[10]. Blaschke, T., Kelly, M., & Merschdorf, H. (2024). 8 Object-Based Image Analysis. *Remote Sensing Handbook, Volume*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

II: Image Processing, Change Detection, GIS, and Spatial Data Analysis, 145.

[11]. Yuan, M., Xia, Y., Dong, H., Chen, Z., Yao, J., Qiu, M., ... & Zhang, L. (2023). Devil is in the queries: advancing mask transformers for real-world medical image segmentation and out-of-distribution localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 23879-23889).

[12]. Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., ... & Gao, M. (2023). Techniques and challenges of image segmentation: A review. *Electronics*, *12*(5), 1199.

[13]. Deng, R., Zhou, M., Huang, Y., & Tu, W. (2023). Hybrid Task Cascade-Based Building Extraction Method in Remote Sensing Imagery. *Remote Sensing*, *15*(20), 4907.

[14]. Arosio, R., Hobley, B., Wheeler, A. J., Sacchetti, F., Conti, L. A., Furey, T., & Lim, A. (2023). Fully convolutional neural networks applied to large-scale marine morphology mapping. *Frontiers in Marine Science*, *10*, 1228867.

[15]. Zhao, X., Zhang, K., Zhang, F., Sun, J., Wan, W., & Zhang, H. (2024). Multi-View Hypergraph Fusion Network for Change Detection in High-Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

[16]. Long, X., Zhuang, W., Xia, M., Hu, K., & Lin, H. (2023). SASiamNet: Self-adaptive Siamese Network for change detection of remote sensing image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

[17]. Yin, M., Chen, Z., & Zhang, C. (2023). A CNN-Transformer Network Combining CBAM for Change Detection in High-Resolution Remote Sensing Images. *Remote Sensing*, *15*(9), 2406.

[18]. Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018, July). Urban change detection for multispectral earth observation using convolutional neural networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2115-2118). IEEE.

[19]. Gomroki, M., Hasanlou, M., & Chanussot, J. (2023). Automatic 3D Multiple Building Change Detection Model based on Encoder-Decoder Network using Highly Unbalanced Remote Sensing Datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

[20]. Vasavi, S., Somagani, H. S., & Sai, Y. (2023). Classification of buildings from VHR satellite images using ensemble of U-Net and ResNet. *The Egyptian Journal of Remote Sensing and Space Sciences*, *26*(4), 937-953.

[21]. Daudt, R. C., Le Saux, B., Boulch, A., & Gousseau, Y. (2018, July). Urban change detection for multispectral earth observation using convolutional neural networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2115-2118). IEEE.