Eurasian J. Sci. Eng., 2025, 12(1), 1-22
https://doi.org/10.23918/eajse.v12i1p1

Published by Tishk International University
Available at: https://eajse.tiu.edu.iq/

RESEARCH ARTICLE Open Access

Urban Building Change Detection Using VHR Satellite Imagery and Deep Learning

Taha K. Azeez [1]*, and Haval A. Sadeq [1]

[1]Department of Geomatics (Surveying) Engineering, College of Engineering, Salahaddin University-Erbil, Kurdistan Region, Iraq.

Article History
Received: 19.01.2026
Revised: 22.05.2026
Accepted: 09.06.2026
Published: 10.06.2026
Communicated by: Prof. Dr. Ayad M. Fadhil Al-Quraishi
*Email address:
tahakamal777@gmail.com
*Corresponding Author

Abstract: Urban building change detection using very high-resolution (VHR) satellite imagery plays an important role in urban planning and monitoring, particularly in rapidly developing areas. This study aims to evaluate the effectiveness of a deep learning-based architecture for detecting urban building changes using very high-resolution satellite imagery.

The analysis focuses on two distinct areas with various types of buildings where typologies are located in Erbil Governorate, Kurdistan Region of Iraq. The dataset consists of bi-temporal satellite images that were acquired by WorldView-2 (2010) and WorldView-3 (2025) with a spatial resolution of 0.5m. The first study area (Korian) covering 0.80km^2, second study area (Peshang) covering 0.88km^2, third study area (Spanish) covering 0.61km^{2}and the fourth study area (Zerin) covering 0.34 km^2. In the experiment, the U-Net-based neural network architecture with a ResNet34 encoder was implemented and trained using a sparse labeling strategy, combined with masked binary cross-entropy loss to consider the problem that arises due to class differences and imitated labelling. The model performance was evaluated against a Random Forest classifier trained on manually engineered spectral features using identical training and testing conditions. The evaluation was conducted using standard performance metrics, including accuracy, F1-score, and Intersection over Union (IoU).

The findings indicate that the deep-learning model steadily outperforms the Random Forest classifier in all study areas. Overall accuracies were shown to be above 95% with great improvements in F1-score and IoU, which indicates improved robustness against class imbalance and improved spatial delineation of changes in the building. However, the model's performance is influenced by the limited availability of training data, which restricts the model's generalization. Furthermore, variants in the model sensor and acquisition condition between the two data sets might lead to inconsistencies in feature detection.

Overall, the study shows the potentiality of using deep learning in changed detection for accurate and scalable urban growth monitoring, while emphasizing on focusing the requirement for larger training and more diverse datasets across different areas for the purpose of generalizing the model.

Keywords: *Urban Building Change Detection; Very-High-Resolution Satellite Imagery; Deep Learning; U-Net; Random Forest.*

1. Introduction

Intensive urbanization has been shown to be one of the most dramatic changes in the global environment of the 21st century that has extensive negative and positive impacts on environmental sustainability, resources, and human well-being [1]. Detecting spatial and temporal changes associated with built environments has become one of the most important applications of remote sensing since it

helps in addressing different fields such as urban planning, disaster monitoring, infrastructure inspection, and environmental management [2]. Monitoring urban expansion, building construction, and changes in land use provide decision-makers with critical data for managing increasingly complicated metropolitan systems.

Traditional techniques of urban change detection have been mainly based on pixel-based and object-based image analysis, using techniques like image differencing, principal component analysis, and supervised classification algorithms, including Support Vector Machines and the Random Forest classifiers [3, 4]. Although these traditional approaches have proven to be effective in specific contexts, they are frequently challenged by the complexity of very high-resolution (VHR) satellite imagery, specifically due to spectral variability, shadowing, and varied building types that can cause serious detection errors [5]. Random Forest classifiers, specifically, have been broadly used to establish change detection based on spectral bands, spectral indices, and temporal difference features, attaining a moderate level of performance for medium-resolution imagery, but having challenges when modeling complex spatial patterns in VHR data [6].

With the rapid advancement of deep learning, remote sensing has shown tremendous progress through advanced automated feature extraction and pattern recognition in satellite images. Conventional neural networks (CNNs), especially architectures such as U-Net, have proven their exceptional performance in semantic segmentation and change detection tasks [7, 8]. Having been originally designed for biomedical image segmentation, the U-Net architecture has been widely adapted to find building footprints and detect urban change, due to its encoder-decoder architecture with skip connections, which successfully preserves spatial details and acquires features at various scales [9, 10]. Further improvement includes residual learning frameworks, including U- Net with ResNet encoders, which have further enhanced performance. These designs solve the issue of vanishing gradients and allow more profound network designs [11, 12]. Deep learning algorithms have a significant advantage over traditional methods for automatically recognizing the differences in satellite images of different resolutions. However, among these approaches, CNN architectures like U-Net are particularly effective for change detection tasks, especially when using high-resolution imagery where a large amount of spatial context is important.

VHR satellite images, such as those from commercial satellites such as WorldView-2 and WorldView-3, have become more accessible in the area of urban monitoring applications. These sensors offer panchromatic imagery with a resolution of around 50 cm and multi-spectral functions with 8-band capacity, which makes them offer a higher level of spectral data that considerably increases the accuracy of change detection [13, 14]. The 50 cm spatial resolution provides the possibility to recognize the individual buildings, infrastructure, and smaller urban structures that play a major role in the proper change detection in densely populated areas. Geometric correction and orthorectification can also be greatly enhanced when the results are combined with auxiliary information, such as Digital Elevation Models (DEMs) generated by other sources, such as the Advanced Land Observing Satellite (ALOS) PALSAR, which is necessary in establishing the positional accuracy needed to compare multitemporal changes [15-17].

The study investigates urban change detection in four villages, which are known as Korian, Peshang, Spanish, and Zerin villages, located in the Erbil area of Kurdistan, Iraq, using bi-temporal WorldView-2(2010) and WorldView-3 (2025) images of 0.5 m resolution. The 15-year temporal span captures a major urban transformation that caused by population growth and economic development in the Region. A comparative research framework is adopted, assessing a deep learning technique based on U-Net architecture plus ResNet34 encoder and a conventional random forest-based classifier implementing inputs based on spectral features and band difference. Deep learning model entails the

use of sparse labeling schemes, masked binary cross-entropy loss, and sliding-window inference to derive probability maps, which are later thresholded to produce a final change detection result. This two-fold strategy allows a rigorous performance assessment of deep learning relative to traditional methods, contributing to the growing body of evidence showing the benefits of neural network-related strategies in complex urban change detection situations in VHR satellite imagery.

Despite the significant progress in urban change detection using both traditional machine learning and deep learning approaches, several challenges remain. Traditional methods, such as Random Forest, primarily depend on handcrafted spectral features and are often affected by spectral variability, noise, and shadows in very high-resolution imagery. As a result, they may produce false detections, especially in complex urban areas.

On the other hand, deep learning models like U-Net have shown strong performance. However, most existing studies assume the availability of fully labeled datasets, which is not always realistic in real-world situations. In practice, labeling large VHR datasets is time-consuming and expensive. In addition, limited studies have explored the use of sparse labeling together with deep learning for urban change detection in complex urban environments.

Moreover, only a few studies provide a fair comparison between deep learning and traditional methods under the same experimental conditions using VHR imagery. In addition, the application of advanced deep learning-based change detection methods in the Kurdistan Region of Iraq is still limited.

Therefore, this study aims to address these research gaps by developing an Early Fusion U-Net model with a ResNet34 encoder that incorporates sparse labeling and a masked loss function. The proposed framework improves robustness against noise and enables effective learning from limited annotated data. A comparative evaluation with a Random Forest classifier is also conducted across multiple study areas with different urban development patterns.

2. Relate work

2.1 Previous Studies on Building Change Detection

Change detection based on remote sensing images has developed from early pixel-based approaches [5, 18]. Although computationally efficient when dealing with medium-resolution imagery, these methods were significantly restricted to very high resolution (VHR) satellite data. Very high resolution (VHR) imagery is so complex that spectral confusion arises between structures, like buildings, roads, and pavement, among others, in the case of urban environments [19, 20].

These challenges were addressed by introducing object-based change detection (OBCD) through segmenting the images into expressive objects before starting the classification [21]. [22] developed the LEVIR-CD dataset, a dataset that is used in building change detection that has become a benchmark, which consisted of 637 bi-temporal image pairs with 0.5 m spatial resolution. [14] showed that the 3D orthorectification of very high-resolution (VHR) imagery significantly decreases the number of false positives produced due to building displacement. Conventional supervised classifier models, such as Random Forest, Support Vector Machines, and Decision Trees, have been frequently applied with manually developed features, including spectral indices and texture measures [4]. However, the necessity for selecting manual features in addition to the limited capability for generalization has led to the shift towards learning-based approaches.

2.2 Advances in Deep Learning Algorithms for Image Analysis

Remote-sensing change detection has been dramatically transformed by deep learning, where hierarchical feature representations eliminate the need for manual feature extraction and depend on automatic learning. The most influential architecture that has emerged is the convolutional neural

network (CNN), specifically the Siamese network, where two parallel branches simultaneously process bi-temporal pairs of images [7].

U-Net has been shown to have effective performance because of its encoder-decoder structure, complemented or supported by skip connections that maintain spatial fidelity at various scales [9]. Originally designed to segment biomedical images, U-Net can encode both fine-scale spatial details and semantic information, which makes it especially effective at detecting building transformations in the environment of large cities [10]. During encode stage, the network iteratively down-samples the input to extract abstract representations of the features; on the other hand, the decoder stage recovers the spatial map to the original resolution using skip connections to re-use high-resolution features.

The inclusion of more powerful encoder backbones has shown a major enhancement in U-Net performance. ResNet introduced residual learning by using skip connections to overcome deep network vanishing gradients [11]. The use of the symmetric design of the U-Net and the residual blocks of ResNet facilitates the utilization of deep features and keeps the gradient flow steady [11, 23]. Attention mechanisms with residual learning are much better at delimiting boundaries and minimizing VHR imagery false positives [12].

More recently, transformer-based architectures have leveraged the self-attention technique to compute long-range dependencies [24]. [1] introduced a continuous urban change detection, which is grounded on the temporal feature refinement, achieving state-of-the-art results. At the same time, SNUNet-CD has also been extended to bi-temporal change detection by adding accuracy via dense interconnections [25]. On the other hand, FIBTNet addresses the challenge of pseudo-changes using bi-temporal feature exchange [26]. Self-supervised learning methods have also minimized the need for supervised data, with [27] using contrastive pretraining to learn from unlabeled remote sensing data.

2.3 Challenges and Limitations in Change Detection

Despite the great advancement, there are a number of challenges that restrict the accuracy of change detection that is based on deep learning. Class imbalance is considered a widespread issue because transformed pixels usually occupy a small part of the image, which makes the models biased towards unchanged predictions [5].

The spurious changes may be observed in a temporal separation of image acquisitions due to differences in radiometric and atmospheric conditions in the image rather than actual surface change. The changes in illumination, phenological, and differences in sensor properties are highly evident in very high resolution (VHR) remote sensing data, and shadows on buildings are characterized by high variability [5]. [28] have suggested uncertainty-sensitive methods based on Bayesian rationalization in quantifying predictive confidence and enhancing model robustness.

Accurately drawing building boundaries based on existing models remains challenging, particularly when it is faced with non-regular geometries or buildings in high-density urban locations [29]. Although mechanisms focusing on boundary regions have been explored, attaining pixel-level accuracy still remains problematic [30].

The model generalization throughout different geographic regions, temporal contexts, and sensor types is another continual challenge. Models trained on datasets for a specific region often tend to be lead for a lower performance when they are applied to new destinations that have different urban characteristics [31]. Although transfer learning has proven to be potentially useful, domain modifications have a strong impact on predictive accuracy. Moreover, the issues related to the high computational demand of the new model pose very practical challenges; the models that can be considered the state of the art require a lot of resources, thus making their use in environments with limited computational capabilities extremely challenging [1]. Finally, the fact that there are no generally accepted evaluation protocols that can sufficiently capture the heterogeneity of the real world hinders the objective comparison of methodological approaches [32].

3. Methodology

In this section, the overall workflow of the proposed methodology is illustrated. It starts with wire data acquisition and then preprocessing, then the deep learning process, which is later followed by the accuracy assessment, residual analysis, and comparison analyze as shown in Figure 1.

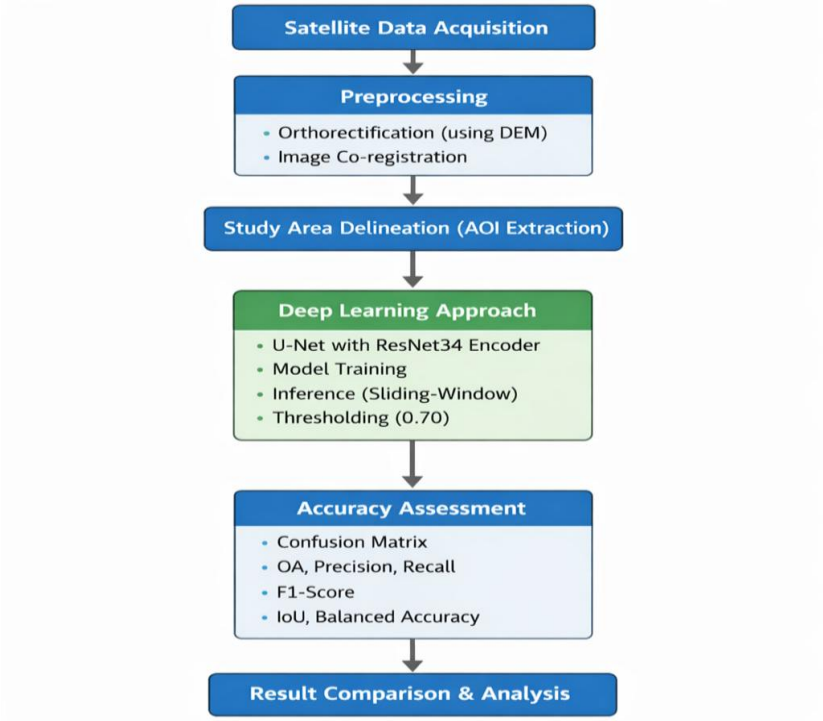

Figure 1: Workflow of the proposed urban change detection methodology.

3.1 Study Area

This research investigates urban change detection in four villages, Korian, Peshang, Spanish, and Zerin, located within the Erbil Governorate of the Kurdistan Region of Iraq. These areas were selected due to the rapid urban expansion and building construction observed in recent years, making them suitable case studies for evaluating change detection methodologies. The locations of the study areas are illustrated in Figure 2.

The selected villages in this paper represent a diverse range of urban development patterns, which include mainly residential areas as well as mixed residential-commercial regions. Such a variety offers a comprehensive testing environment to compare the effectiveness of the deep learning-based algorithms for urban change detection. Besides the planned residential areas, a fourth study area (Zerin) was added to introduce greater complexity. Compared to the other locations, this area contains a combination of new and old structures, resulting in a heterogeneous urban structure. This combination of temporal and structural variations increases the difficulty of the change detection task and provides a more realistic environment for evaluating the effectiveness of the proposed model under varying urban conditions. Moreover, all the study areas include shadow effects and spectral variability, which introduce additional challenges in the accurate building change detection. This allows for a more comprehensive evaluation of the model's robustness under realistic urban conditions.

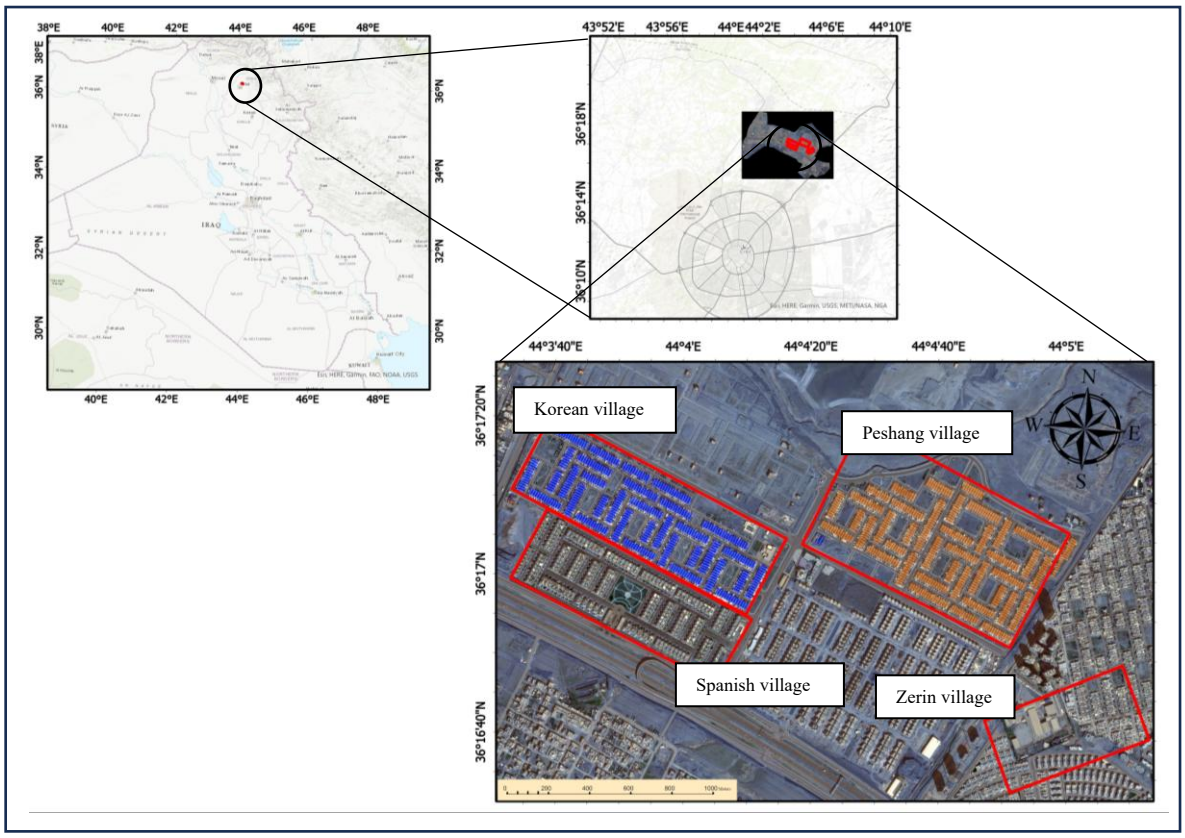

Figure 2: The location of the study areas (Korean, Peshang, Spanish, and Zerin villages) on the VHR satellite imagery.

3.2 Data Collection and Preprocessing

3.2.1 Satellite Imagery Acquisition

The four study areas were covered by very high-resolution (VHR) satellite imagery that was supplied by Space Imaging Middle East (SIME).

Bi -temporal satellite images were collected from WorldView-2 (WV-2) on July 4, 2010, and WorldView -3 (WV-3) on February 10, 2025, with a spatial resolution of 0.5 m. The images were captured in different seasons (summer and winter), allowing evaluation under varying illumination conditions. The imagery was provided in the form of standard orthorectified products in GeoTIFF format with a 16-bit radiometric resolution. The four spectral bands, such as red, green, blue, and near-infrared, were available for each scene and were pansharpened to achieve a 0.5 m spatial resolution. This study utilized four spectral bands (red, green, blue, and near-infrared) rather than the complete eight-band multispectral data to ensure consistency between the WorldView-2 and the WorldView-3 datasets. Additionally, the use of pan-sharpened 0.5 m imagery is essential for capturing fine spatial details required for accurate building change detection. These sensors have a capability of capturing spectral data, which helps in the extraction of spectral indices and features, which are critical in the differentiation of the building from other land-cover categories and in identifying changes in the urban environment over time.

3.3 Preprocessing Workflow

Satellite image preprocessing was done using ERDAS Imagine and with a standardized workflow that is used to ensure geometric and radiometric consistency of bi-temporal datasets.

In order to attain high geometric accuracy in preprocessing, a digital elevation model (DEM) with a spatial resolution of 12.5 m was obtained from the Alaska Satellite Facility (ASF) and used in Orthorectification process.

The 12.5 m resolution DEM was considered sufficient for this study, as it was not used for terrain analysis but was primarily used for orthorectification and geometric correction. Higher-resolution elevation data, including those generated by UAVs, can provide more detailed topographic information; however, such data are not always readily available and were not feasible for this study. Consequently, the selected DEM provides an appropriate balance between accuracy and data availability for the intended application. The ordered satellite imagery provided with rational polynomial coefficients (RPCs) that were related to Worldview imagery, which was also used during the orthorectification process. This process minimized terrain-related distortions and ensured that the same ground locations were represented by the corresponding pixels of various acquisition dates, which is a prerequisite for reliable urban change detection.

Subsequently, image co-registration was used to enable accurate alignment of the orthorectified images. For accurate co-registration, subpixel alignments between images have been applied by using automated image-matching algorithms. Even a minor co-registration error can lead to a false change detection, particularly in a highly urbanized environment.

Following geometric preprocessing, the imagery was imported into ArcGIS Pro for quality control and area of interest (AOI) extraction. Visual inspection as well as basic quantitative checks were conducted to verify the accuracy of orthorectification and co-registration. The selected study areas (Korian, Peshang, Spanish, and Zerin villages) were defined as polygon features and used to obtain spatially consistent image subsets of the full satellite scenes. These AOI subsets were subsequently exported and prepared for further analyzing by change-detection methods.

3.4 Training Data and Ground Truth

Training data that has been used in the deep learning model was created by manual digitization of building changes using ArcGIS Pro. A sparse labeling approach was used, in which polygonal areas were used to outline the change and no-change areas. This strategy considerably reduced manual annotation effort while maintaining sufficient spatial variability, thus allowing the model to learn discriminative features effectively. The annotated polygons were then converted into binary raster masks with a spatial resolution of 0.5m similar to the original input very-high-resolution (VHR) satellite images. In these masks, the pixel labeled as 1 means that there is change and pixel labeled as 0 stands for no change.

Each of the study areas (Korian, Peshang, Spanish, and Zerin villages) had independent ground-truth datasets they are used only to evaluate the model. These reference datasets provided reliable data about the real building alterations and made it possible to compare the deep learning system and the Random Forest classifier (which was used in the result comparison) objectively.

Both training and ground-truth datasets were subsequently used as input to the deep learning structure described in the section 3.3, where the U-Net architecture with ResNet34 encoder is explained.

3.5 Deep Learning Model Architecture

3.5.1 U-Net with ResNet34 Encoder

The deep learning architecture adopted in this study is based on the U-Net architecture with a ResNet34 encoder backbone. The architecture of the proposed Early Fusion U-Net model is illustrated in Figure3.

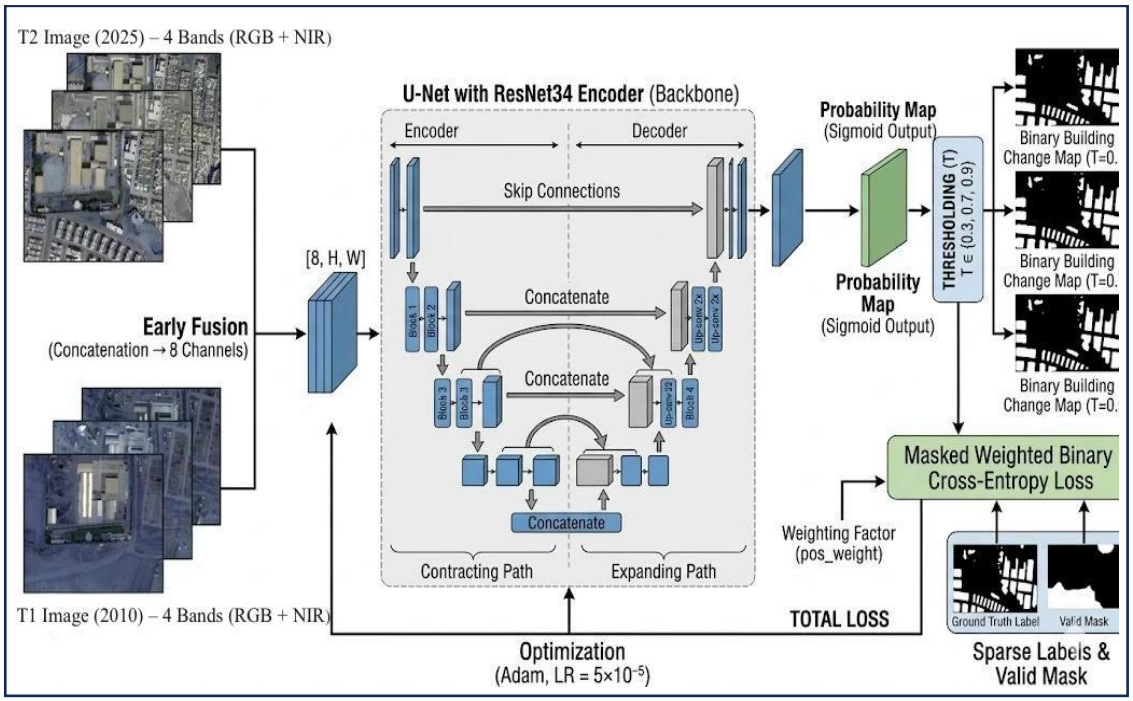

Figure 3: Architecture of the proposed Early Fusion U-Net model with a ResNet34 encoder for urban building change detection. Bi-temporal satellite images (2010 and 2025), each consisting of four spectral bands (RGB and NIR), are concatenated into an 8-channel input and processed through an encoder–decoder network with skip connections. The model generates a probability map using a sigmoid activation function, followed by thresholding to produce a binary building change map. A masked weighted binary cross-entropy loss is employed to address class imbalance and sparse labeling conditions.

Originally designed to perform biomedical image segmentation, the U-Net architecture has been widely implemented in remote sensing applications due to its encoder-decoder structure with skip connections that lead to effectively balancing precise spatial localization and the preservation of contextual features.

In this architectural design, the standard U-Net encoder is replaced with ResNet34, which is a 34-layer residual network that is specifically designed to learn robust feature representations by using identity-based residual connections. These residual connections are effective in reducing the vanishing gradient problem, allowing the model to be trained in deeper networks and increasing convergence stability. Lower network layers focus on the extraction of the elementary spatial features, such as edges and textures, and higher layers incorporate more abstract and semantically significant features, which are relevant to urban morphology and building structure.

The decoder mirrors the encoder structure; progressively, the feature maps are upsampled to reconstruct the output map at the original spatial resolution. By combining high-level semantic features with low-level spatial features, which were acquired in the respective encoder layers, are used to create skip connections. This process is used to maintain the spatial accuracy that would have been lost during downsampling. This architecture is particularly designed for high-resolution urban imagery, where more accurate delineation of boundaries is necessary to obtain reliable results in building-change detection.

A sigmoid activation function is used at the output layer to generate a pixel-level probability map, where every pixel value indicates the likelihood of change within the continuous range [0,1].

3.6 Training Strategy

The U-Net model was trained using a sparse labeling strategy, in which manually digitized samples were used to direct the learning process. To reduce the effect of the class imbalance caused by the difference between the altered and unaltered pixels in the change detection, a masked binary cross-entropy (BCE) loss function was used.

A masked weighted binary cross-entropy loss was used to handle class imbalance and sparse labeling conditions. Although alternative loss functions such as Dice loss and Focal loss are commonly used in image segmentation, binary cross-entropy was selected due to its stability and suitability for the given problem. This loss formulation focuses on optimization on annotated pixels, thus enabling effective learning in the case of limited supervision and overcoming the bias toward the majority class with the use of the corresponding weight of the classes.

The model was optimized by using the Adam optimizer; an initial learning rate was determined experimentally, and the optimization was accelerated with the use of a GPU. The model was trained for 90 epochs using a batch size of 2 and an initial learning rate of 5×10^{-5}. No weight decay was applied during training. All experiments were conducted in a CUDA-enabled GPU environment. Data augmentation methods were applied to enhance generalization and reduce overfitting. This method included flipping, rotational transformations, and luminance adjustments. Training was continued until convergence was achieved, which was identified by the loss of validation stabilization.

3.7 Inference and Post-Processing

The probability maps were then converted into binary change maps using thresholding. An empirical analysis was systematically performed by changing the threshold value between 0.00 and 1.00 in steps of 0.02 and evaluating the performance of the resultant values against ground-truth values. The threshold of 0.70 was selected as optimal, providing a trade-off between accuracy in change detection and the reduction of false positives. This threshold was applied in all the study regions, resulting in producing the final change detection products.

3.7.1 Random Forest Classification

To provide a conventional basis of comparison, a Random Forest (RF) was implemented to be applied in comparative analysis.

Random Forest is an ensemble learning technique that builds based on a vast number of decision trees at the training stage and sums up the results of each of the trees to provide a final classification result. Owing to its robustness to noise, its ability to be operated in high-dimensional feature spaces, and its steady functioning implementation on a limited dataset, Random Forest has been widely used in the field of remote sensing, especially in land-cover classification and change detection.

The random forest classification algorithm was used in this study to process bi-temporal, very high-resolution satellite images, with the use of manually chosen spectral features. The features that were selected included spectral band differences, spectral ratios, and normalized difference indices computed between the two acquisition dates; these measures are widely used to describe changes in surface reflectance spectral values as a result of land use and land cover change.

The random forest (RF) model was also trained in a supervised way with the same labeled data that were used to train the deep learning model, thus providing a fair and consistent platform to compare them. Predictions of pixel-level change were done in each region under study (Korian, Peshang, Spanish, and Zerin villages), and the results were compared to independent ground-truth data using the same accuracy measures. This experimental design also made the objective comparison of the traditional Random Forest technique with the new deep learning technique, thus explaining the difference in the detection performance under the same data and assessment conditions.

The Random Forest model was implemented using 100 trees (n_estimators = 100) with no restriction on maximum depth (max_depth = None). The Gini impurity criterion was used for node splitting, and the number of features considered at each split was set to the square root of the total number of features (max_features = "sqrt"). Other parameters were kept at their default values.

Feature importance analysis was conducted based on the trained model to identify the most influential variables contributing to change detection.

No hyperparameter tuning or cross-validation was performed, as the aim of this study was to provide a consistent baseline comparison with the deep learning model. All models were trained and evaluated under the same data and experimental conditions to ensure a fair comparison.

3.7.2 Evaluation Metrics

To perform an in-depth evaluation of the results of both deep learning and Random Forest models, several accuracy measures were obtained based on the confusion matrices that were built on the basis of the comparison of the predicted binary change maps and the corresponding ground-truth information. Every confusion matrix would consist of four elements, namely: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Based on these aspects, the following evaluation metrics have been computed [32, 33]:

Overall Accuracy (OA) measures the proportion of correctly classified pixels:

(1) $$\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

In the literature, precision is referred to as producer accuracy, and it is a measure of the reliability of changes that are detected.

(2) $$\text{Precision} = \frac{TP}{TP+FP}$$

Recall, or sensitivity, as it is also known as producer accuracy, measures the percentage of true changes that are recognized.

(3) $$\text{Recall} = \frac{TP}{TP+FN}$$

Specificity evaluates the correct identification of unchanged pixels:

(4) $$\text{Specificity} = \frac{TN}{TN+FP}$$

The F1 - score, which is the harmonic mean of precision and recall, is a balanced metric for evaluating the performance of detection:

(5) $$\text{F1-Score} = \frac{2\times Precision\times Recall}{Precision+Recall} = \frac{2\times TP}{2\times TP+FP+FN}$$

The Jaccard Index is also called intersection over Union (IoU), which measures the spatial concordance between predicted and reference change areas:

(6) $$\text{IoU} = \frac{TP}{TP+FP+FN}$$

Balanced Accuracy accounts for class imbalance by averaging recall and specificity:

(7) $$\text{Balanced Accuracy} = \frac{Recall+Specificity}{2}$$

Each evaluation metric was calculated separately for each study area (i.e., Korian, Peshang, Spanish, and Zerin villages) to determine the strength, consistency, and applicability of both methods across dissimilar urban settings.

4. Results

This section presents the qualitative and quantitative results obtained from the proposed deep learning-based building change detection framework and compares them with the Random Forest classifier across the four study areas.

4.1 Visualization of Building Change Detection Results

The following subsection will provide a visual comparison of the building change detection results of the deep learning model and the Random Forest classifier in the four study areas. The visual analysis will help to better understand the way each method detects building changes and how the identified changes in the buildings relate to the actual urban patterns.

4.2 Ground-Truth Reference Maps

The ground-truth maps are maps that outline the true building transformations that occurred between the two-time snapshots and were produced based on the verified building footprints. These maps are an assessment framework reference for the performance of change-detection methods. Figure 4 shows the spatial distribution of altered and unchanged areas in the four villages.

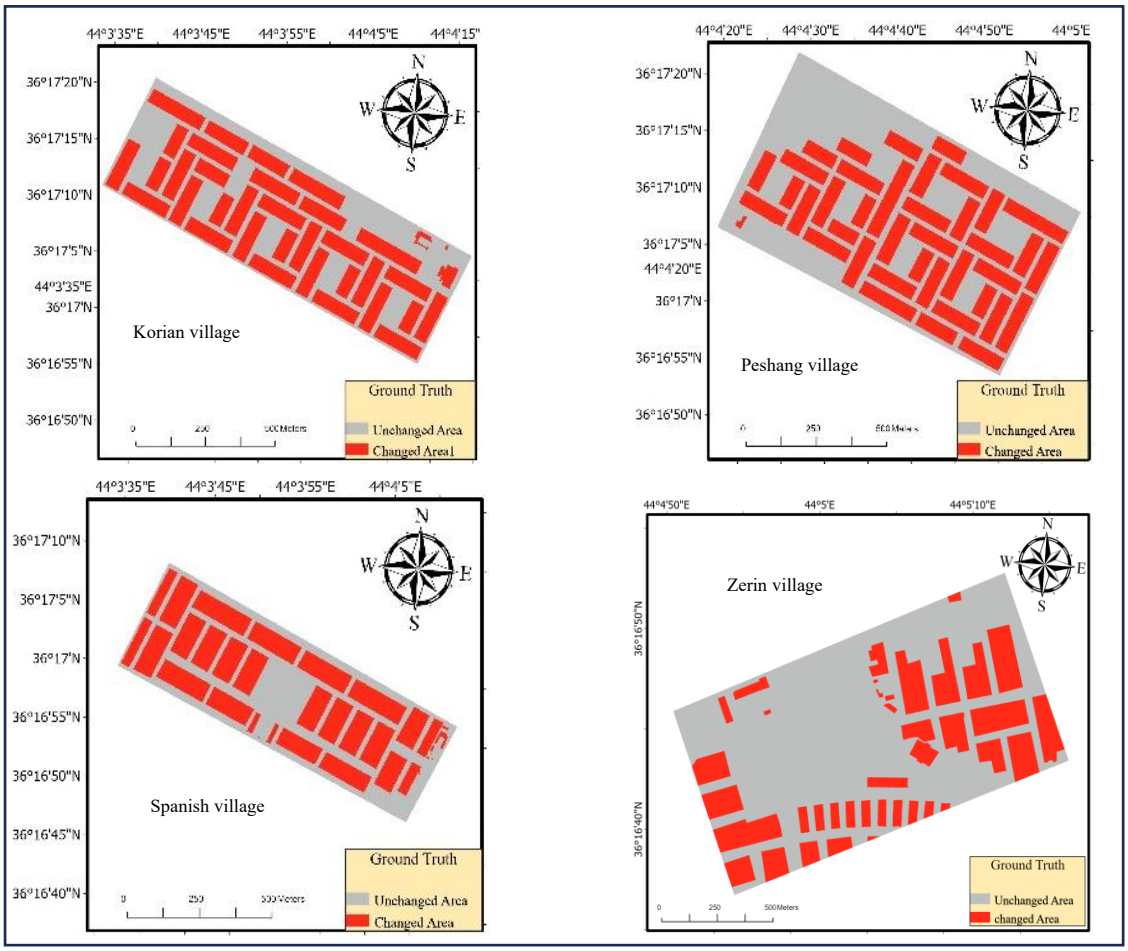

Figure 4: illustrates the ground-truth building change maps of the villages of Korian, Peshang, Spanish, and Zerin that were obtained based on validated building footprints. In these maps, the pixel values of the change areas are denoted with the value of 1, and the pixel values of the unchanged

areas are denoted with the value of 0. These reference maps make up the authoritative foundation on which quantitative and qualitative evaluations are conducted.

4.3 Deep Learning Detection Results

The outputs of the deep learning model indicate evident and precise changes in building forms with lower noise in all the study areas. Large and small buildings are correctly recognized, and the spatial patterns that have been obtained are very close to the reference maps. Figure 5 illustrates these findings

Figure 5: illustrates the building change detection results produced by the deep learning model. The predicted change maps have well-defined building footprints, low levels of noise, and a high degree of spatial coherence in all four study areas. Large residential complexes and smaller individual buildings are outlined with accuracy and are very close to the spatial patterns that exist in the reference maps.

4.4 Traditional Random Forest Results

Compared to the outcomes of deep learning techniques, change detection results based on the Random Forest classifier exhibit higher noise and poorer delineation of building boundaries. In particular, areas

with shadows or intricate surface geometry are more prone to false change detections. Figure 6 illustrates these phenomena.

Figure 6: shows the change detection results generated using the Random Forest classifier. Compared to the outputs of the deep learning models, the resulting maps are characterized by an increased level of salt-and-pepper noise, reduced fidelity of building edges, and a higher number of false positives, in particular, in spectrally richer or shadowed areas.

4.5 Error Map Analysis

The error maps were generated to further assess the performance of the model by comparing the predicted results with the ground-truth data. These maps highlight the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), providing a clear visualization of the classification errors across the study areas. As shown in Figure 7, the deep learning model produces more spatially consistent results with fewer false detections, whereas the Random Forest method exhibits higher levels of noise and misclassification.

Figure 7: Error maps for deep learning (left) and Random Forest (right) across the study areas, showing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

4.6 Quantitative Analysis of Model Performance

4.6.1 Overall Performance Comparison

The quantitative analysis reveals that the proposed deep learning method outperformed the traditional Random Forest (RF) classifier in all four research spheres. Table 1 provides a detailed comparison of the key accuracy measures obtained with the confusion matrices, such as Overall Accuracy, Precision, Recall, F1 -score, Intersection over Union (IoU), and Balanced Accuracy.

Table 1: Comparative Accuracy metrics of Deep Learning and traditional Approaches to the Random Forests under all the study areas.

Study Area	Method	Accuracy	Precision	Recall	F1-Score	IoU	Balanced Accuracy
Korian	Random Forest	88.26%	83.49%	95.59%	89.13%	80.40%	88.20%
Korian	Deep Learning	95.63%	94.52%	96.94%	95.72%	91.78%	95.62%
Peshang	Random Forest	88.99%	81.27%	96.45%	88.21%	78.91%	89.94%
Peshang	Deep Learning	96.23%	94.01%	97.36%	95.66%	91.68%	96.37%
Spanish	Random Forest	87.09%	88.23%	87.91%	88.07%	78.68%	87.01%
Spanish	Deep Learning	95.60%	93.99%	98.15%	96.02%	92.35%	95.36%
Zerin	Random Forest	83.98%	64.79%	97.61%	77.80%	63.80%	94.10%
Zerin	Deep Learning	96.48%	92.78%	95.14%	93.99%	88.95%	96.05%

To assess the statistical reliability of the reported accuracy values, 95% confidence intervals were computed based on the binomial distribution of correctly classified pixels. The results indicate that the deep learning model achieves an overall accuracy of 95.63% ± 0.05% in Korian, 96.23% ± 0.04% in Peshang, 95.60% ± 0.05% in Spanish, and 96.48% ± 0.04% in Zerin. These narrow confidence intervals confirm the robustness and statistical stability of the proposed model across different urban environments.

In the four residential compounds, the deep-learning model had an overall accuracy of 95.63% (Korian), 96.23% (Peshang), 95.60% (Spanish), and 96.48% (Zerin), that is, 7.37%, 7.24%, 8.51%, and 12.5% higher than the Random Forest classifier had. These advantages give results indicating the higher ability of deep-learning techniques in capturing urban-building variations in a heterogeneous spatial arrangement and in spectral conditions.

The identified improvements in both F1-score and IoU further support the ability of the deep learning model to address the issue of imbalance in the classes and maintain reliable spatial matching with ground-truth information.

4.6.2 Confusion Matrix Analysis

To provide deeper insight into how the classification behavior is characterized across metrics, the confusion matrices of the two methods are presented in Table 2, in all areas of the study.

Table 2: Confusion matrices for deep learning and Random Forest methods across all study areas.

Area	Method	True Negative	False Positive	False Negative	True Positive	Total Pixels
Korian	Random Forest	559,534	132,887	30,975	671,989	1,395,385
Korian	Deep Learning	653,515	39,523	21,493	681,471	1,396,002
Peshang	Random Forest	904,002	179,511	28,659	778,696	1,890,868
Peshang	Deep Learning	1,034,643	50,087	21,330	786,719	1,892,779
Spanish	Random Forest	424,318	68,444	70,529	512,837	1,076,128
Spanish	Deep Learning	456,761	36,622	10,783	572,583	1,076,749
Zerin	Random Forest	394184	108796	4904	200142	708026
Zerin	Deep Learning	488004	14976	9950	195096	708026

The confusion matrix analysis shows similar patterns. The deep learning methodology significantly reduced the number of false-positive identifications across the surveyed areas, with maximum impact recorded in Peshang, where the number of false identifications reduced to 50,087 pixels as opposed to 179,511 pixels. At the same time, the number of true positives was either preserved or even better yet improving the high-quality detection of genuine building changes without compromising sensitivity.

Moreover, the deep learning model revealed greater consistency in different urban settings. This is supported by the fact that the standard deviation of the overall accuracy in the four areas of study was less than it was with the Random Forest classifier, which consequently indicates that the proposed approach has a better generalization ability in a heterogeneous urban setting.

4.6.3 Precision, Recall, and F1-Score Analysis

Precision and recall are critical metrics for evaluating change-detection performance, as they describe the reliability of detected changes and the sensitivity to actual changes, respectively. Table 1 shows that the deep-learning approach has yielded significantly greater values of precision at all the studied areas, reaching 94.52% in Korian, 94.01% in Peshang, and 93.99% in Spanish, and 92.78 % in Zerin compared with 83.49%, 81.27%, 88.23%, and 64.79 obtained by the Random Forest classifier.

The recall values for the deep learning model were consistently high, with 96.94%, 97.36%, 98.15%, and 95.14% in Korian, Peshang, Spanish, and Zerin, respectively. Such findings illustrate a high level of sensitivity to any building changes and low error rates of omission, thereby demonstrating the model's ability to identify the real structural transformations in complex and heterogeneous urban areas.

The trade-off between precision and recall, which is quantified by the F1-score, further highlights the strength of the deep-learning method. It was found that there were improvements of 6.59%, 7.45%, 7.95%, and 16.19% compared to the traditional methods in Korian, Peshang, Spanish, and Zerin, respectively, demonstrating that the proposed model simultaneously provides high detection reliability and strong sensitivity.

4.6.4 Intersection over Union (IoU) Evaluation

Intersection over Union (IoU) provides a rigorous assessment of the spatial agreement between predicted change areas and the reference ground truth. The deep learning model achieved IoU values of 91.78% (Korian village), 91.68% (Peshang village), 92.35% (Spanish village), and 88.95% (Zerin village), significantly higher than the Random Forest method, 80.40%, 78.91%, 78.68%, and 63.80%, respectively.

The comparison of the results revealed that the deep learning methodology provides better building boundaries and generates change maps with higher spatial coherence, as reflected by IoU improvements of 11.38%, 12.77%, 13.67%, and 25.15 across all study areas. The current result is especially important for urban applications, where the correct delineation of the building footprints becomes essential for planning, monitoring, and management purposes.

4.6.5 Statistical Significance Analysis

A McNemar test was performed on the Zerin study area using pixel-level predictions from both the deep learning model and the Random Forest classifier. Zerin was selected as a representative area for statistical evaluation. The test produced a p-value of 0.0 ($p < 0.05$), which shows that the difference between the two methods is statistically significant. This means that the better performance of the deep learning model is not due to chance, but reflects a real improvement in change detection results.

4.7 Benchmark Comparison

The proposed model achieved an overall accuracy of approximately 96% and an F1-score of approximately 95%, which compares favorably with results reported in previous deep learning-based change detection studies. Early convolutional methods such as FC-EF and FC-Siam-diff [7] reported F1-scores in the range of 74–80%, while UCDNet [8] achieved approximately 82–85%. More recent architectures, including STANet [22], SNUNet-CD [25], and ChangeFormer [24], reported F1-scores between 87% and 90% on benchmark datasets such as LEVIR-CD. Advanced models applied to very high-resolution imagery, such as DAttResU-Net [12], have achieved F1-scores of approximately 91–93%.

In comparison, the proposed model achieved an F1-score of approximately 95%, placing it in the upper range of reported performance and, in some cases, exceeding existing methods. Although direct comparison is limited due to differences in dataset characteristics, sensor type, and spatial resolution, the results indicate that the proposed approach performs competitively with current state-of-the-art methods for building change detection.

5. Discussion

5.1 Overall Discussion of Model Performance

The experimental results have clearly shown that the deep learning-based change detection model is superior when compared to the traditional Random Forest framework in all the explored fields of study. The steady high rates of Overall Accuracy, F1 -score, and Intersection over Union (IoU) are a testament to the increased performance of the deep learning model in outlining the changes of complex urban buildings in the heterogeneous environment.

The significant increase in F1 -score highlights the ability of the model to balance accuracy and recall with high class imbalance, where the same pixels occupy a disproportionately large part of the image. This result implies that the given model is rather efficient in the reduction of bias against the majority category, which is a common problem of conventional pixel-based classifiers that are frequently used in urban change classification.

Similarly, the continuously high scores of the IoU are in support of the fact that the deep learning model not only detects changes with high precision but also provides building boundaries with high spatial fidelity. Urban uses cannot avoid such spatial accuracy, with proper extraction of the footprint being at the basis of planning, monitoring, and management of infrastructure.

The better performance of the proposed model is also supported by statistical analysis, including confidence intervals and the McNemar test. These results show that the difference between the models is statistically significant and not due to chance.

These findings are consistent with recent studies in the literature, where deep learning models have demonstrated superior performance in urban change detection tasks.

5.2 Comparison with the Traditional Random Forest Method

The comparative analysis highlights key structural differences between deep learning and traditional machine learning approaches. Unlike the Random Forest classifier, which relies on manually engineered spectral descriptors, the deep learning model automatically builds hierarchical spatial spectral representations that are directly learnable from the input data. As a result, the deep-learning architecture demonstrates a higher ability to distinguish between real building modifications and spectrally introduced distortions that can be explained by the presence of shadows, changes in lighting, or the change of time of the year.

This conclusion is further supported by the reduction in false positives, as shown by the analysis of the confusion matrix. Interestingly, the deep learning model is continuously able to remove non-useful detections and preserve, or even improve, the actual positive rates, and, therefore, a stronger discrimination between the true structural change and the non-significant background variability.

5.3 Generalization across Urban Contexts

An important finding of this study is the high generalization ability portrayed by the proposed deep-learning model in various urban settings. Although the building density, the spatial distribution, and the building schemes of the villages: Korian, Peshang, Spanish, and Zerin differ, the model was able to maintain a stable performance, as it was exhibited by a low variability in the measures of accuracy.

The reduced standard deviation of overall accuracy in all areas of the study, compared to the Random Forest standard, supports the fact that such a deep-learning architecture is less sensitive to local urban morphology and spectral heterogeneity. This robustness makes the framework particularly beneficial to operational urban monitoring systems, which are required to perform dependably over a range of different and fast-changing conditions.

The improved accuracy of the deep learning model in detecting buildings and its better spatial coherence result in real-world utility in urban management. The high-quality building-change maps can be used to inform a range of applications such as detecting informal settlements, monitoring urban sprawl, updating cadastral databases, and supporting the assessment of damages after a disaster.

Since the model can run well on sparse training data, the proposed framework can also be applied to areas with limited availability of ground truth data and thus can be considered a scalable one for large-scale urban monitoring.

Despite its strong performance, the proposed approach has certain limitations. Radiometric differences between acquisition dates, particularly those related to seasonal vegetation changes or shadow effects, may still introduce minor false detections. Incorporating additional data sources, such as multi-temporal SAR imagery or digital surface models, could further improve robustness.

Moreover, the current framework is based on bi-temporal analysis. The model could be further developed to use multi-temporal time series to better distinguish between long-term structural changes

and short-term changes. It is also possible that future studies will examine the solution of transfer learning to determine the extent to which the learned model can be applied to other geographical regions with different urban features.

Overall, the empirical data findings reveal that deep learning-based change detection algorithms have more accurate, spatially consistent, and generalizable results than traditional methods of supervised machine learning when utilized to detect change in urban buildings. These findings support the effectiveness of the deep learning models in promoting reliable and scalable urban development monitoring in rapidly changing places.

6. Conclusion

This study confirms that the deep learning-based urban building change detection method performs considerably better given the extremely high-resolution satellite images as compared to the Random Forest supervised machine learning methods. The experimental results reveal that the overall accuracy, F1-score, and Intersection over Union (IoU) consistently showed improvement in all areas of the study, which demonstrates to be more robust when there is an imbalance in the classes and better spatial representation of building changes.

The results also indicate that the developed deep learning model can generate spatially consistent change maps that have distinct building outlines, which are necessary for urban analysis. Besides, the model has a high generalization ability across various urban environments, even when the density, layout, and development patterns of buildings differ.

With respect to the applied perspective, the suggested method provides a scalable and efficient solution to the urban monitoring application, such as urban expansion evaluation, identification of informal settlements, cadastral renovation, and post-disaster damage estimation. Overall, the current study confirms that deep learning-based models of change detection have a high potential to ensure reliable, precise, and massive urban monitoring in rapidly expanding areas, which will supply valuable geospatial data to support the creation of evidence-based urban planning and management in the regions that exhibit rapid growing changes.

Conflict of Interest

The authors have no conflicts of interest to declare

Acknowledgment

The authors would like to sincerely thank Salahaddin University-Erbil for the academic environment and facilities provided, which contributed to the successful completion of this study.

Authors contribution

Taha K. Aziz collected and analyzed the data, carried out the experiments, and prepared the manuscript. Haval A. Sadeq supervised the research, provided guidance, and reviewed and edited the manuscript.

Use of AI tool declaration

The authors declare that any AI tools used in the preparation of this manuscript were limited to language and readability improvement only and were not used to generate scientific content, data, analyses, or conclusions, with full responsibility retained by the authors.

References

[1] Hafner S, Fang H, Azizpour H, Ban Y. Continuous urban change detection from satellite image time series with temporal feature refinement and multi-task integration. IEEE Transactions on Geoscience and Remote Sensing. 2025 Jun 11. https://doi.org/10.48550/arXiv.2406.17458

[2] Jiang H, Peng M, Zhong Y, Xie H, Hao Z, Lin J, Ma X, Hu X. A survey on deep learning-based change detection from high-resolution remote sensing images. Remote Sensing. 2022 Mar 23;14(7):1552. https://doi.org/10.3390/rs14071552

[3] Asokan A, Anitha JJ. Change detection techniques for remote sensing applications: A survey. Earth Science Informatics. 2019 Jun 1;12(2):143-60. https://doi.org/10.1007/s12145-019-00380-5

[4] Rynkiewicz A, Hościło A, Aune-Lundberg L, Nilsen AB, Lewandowska A. Detection and quantification of vegetation losses with Sentinel-2 images using bi-temporal analysis of spectral indices and a transferable random forest model. Remote Sensing. 2025 Mar 11;17(6):979. https://doi.org/10.3390/rs17060979

[5] Shafique A, Cao G, Khan Z, Asad M, Aslam M. Deep learning-based change detection in remote sensing images: A review. Remote Sensing. 2022 Feb 11;14(4):871. https://doi.org/10.3390/rs14040871

[6] Cohen WB, Healey SP, Yang Z, Zhu Z, Gorelick N. Diversity of algorithm and spectral band inputs improves Landsat monitoring of forest disturbance. Remote Sensing. 2020 May 23;12(10):1673. https://doi.org/10.3390/rs12101673

[7] Daudt RC, Le Saux B, Boulch A, Gousseau Y. Urban change detection for multispectral earth observation using convolutional neural networks. InIGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium 2018 Jul 22 (pp. 2115-2118). Ieee. https://doi.org/10.48550/arXiv.1810.08468

[8] Basavaraju KS, Sravya N, Lal S, Nalini J, Reddy CS, Dell'Acqua F. UCDNet: A deep learning model for urban change detection from bi-temporal multispectral Sentinel-2 satellite images. IEEE Transactions on Geoscience and Remote Sensing. 2022 Mar 23; 60:1-0. https://doi.org/10.1109/TGRS.2022.3161337

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2015 Oct 5 (pp. 234-241). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

[10] Alsabhan W, Alotaiby T, Dudin B. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. Computational Intelligence and Neuroscience. 2022;2022(1):4831223. https://doi.org/10.1155/2022/4831223

[11] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016 (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

[12] Khankeshizadeh E, Mohammadzadeh A, Mohsenifar A, Moghimi A, Pirasteh S, Feng S, Hu K, Li J. Building detection in VHR remote sensing images using a novel dual attention residual-based U-Net (DAttResU-Net): An application to generating building change maps. Remote Sensing Applications: Society and Environment. 2024 Nov 1; 36:101336. https://doi.org/10.1016/j.rsase.2024.101336

[13] Maxar Technologies. WorldView satellite constellation [Internet]. Westminster (CO): Maxar Technologies; c2024 [cited 2025 Jun 9]. Available from: https://www.maxar.com/constellation

[14] Rizeei HM, Pradhan B. Urban mapping accuracy enhancement in high-rise built-up areas deployed by 3D-orthorectification correction from WorldView-3 and LiDAR imagery. Remote Sensing. 2019 Mar 22;11(6):692. https://doi.org/10.3390/rs11060692

[15] Alaska Satellite Facility. SAR data archive [Internet]. Fairbanks (AK): Alaska Satellite Facility; c2015 [cited 2025 Jun 9]. Available from: https://asf.alaska.edu

[16] Alganci U, Besol B, Sertel E. Accuracy assessment of different digital surface models. ISPRS International Journal of Geo-Information. 2018 Mar 15;7(3):114. https://doi.org/10.3390/ijgi7030114

[17] Abdullah SH, Sadeq HA, Salih DM. 3D Buildings Change Detection from Aerial and Satellite Stereo Imagery Using the Kullback–Leibler Divergence Algorithm. Wasit Journal of Engineering Sciences. 2022;10(4):13-26. https://doi.org/10.31185/ejuow.Vol10.Iss4.454
[18] Lu D, Mausel P, Brondizio E, Moran E. Change detection techniques. International journal of remote sensing. 2004 Jun 1;25(12):2365-401. https://doi.org/10.1080/0143116031000139863
[19] Myint SW, Gober P, Brazel A, Grossman-Clarke S, Weng Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote sensing of the environment. 2011 May 15;115(5):1145-61. https://doi.org/10.1016/j.rse.2010.12.017
[20] Salehi B, Zhang Y, Zhong M, Dey V. Object-based classification of urban areas using VHR imagery and height points ancillary data. Remote Sensing. 2012 Aug 3;4(8):2256-76. https://doi.org/10.3390/rs4082256
[21] Hussain M, Chen D, Cheng A, Wei H, Stanley D. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS Journal of photogrammetry and remote sensing. 2013 Jun 1; 80:91-106. https://doi.org/10.1016/j.isprsjprs.2013.03.006
[22] Chen H, Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. Remote sensing. 2020 May 22;12(10):1662. https://doi.org/10.3390/rs12101662
[23] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-Net. IEEE Geoscience and Remote Sensing Letters. 2018 Mar 8;15(5):749-53. https://doi.org/10.1109/LGRS.2018.2802944
[24] Chen H, Qi Z, Shi Z. Remote sensing image change detection with transformers. IEEE Transactions on Geoscience and Remote Sensing. 2021 Jul 20; 60:1-4. https://doi.org/10.1109/TGRS.2021.3095166
[25] Fang S, Li K, Shao J, Li Z. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. IEEE Geoscience and Remote Sensing Letters. 2021 Feb 17; 19:1-5. https://doi.org/10.1109/LGRS.2021.3056416
[26] Wang J, Lin T, Zhang C, Peng J. FIBTNet: Building Change Detection for Remote Sensing Images Using Feature Interactive Bi-Temporal Network. Computers, Materials & Continua. 2024 Sep 1;80(3). https://doi.org/10.32604/cmc.2024.053206
[27] Feng W, Guan F, Tu J, Sun C, Xu W. Detection of changes in buildings in remote sensing images via self-supervised contrastive pre-training and historical geographic information system vector maps. Remote Sensing. 2023 Dec 8;15(24):5670. https://doi.org/10.3390/rs15245670
[28] Li J, He W, Li Z, Guo Y, Zhang H. Overcoming the uncertainty challenges in detecting building changes from remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing. 2025 Feb 1; 220:1-7. https://doi.org/10.1016/j.isprsjprs.2024.11.017
[29] Guo H, Du B, Zhang L, Su X. A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing. 2022 Jan 1; 183:240-52. https://doi.org/10.1016/j.isprsjprs.2021.11.005
[30] Lin H, Hao M, Luo W, Yu H, Zheng N. BEARNet: A novel building edge-aware refined network for building extraction from high-resolution remote sensing images. IEEE Geoscience and Remote Sensing Letters. 2023 May 2; 20:1-5. https://doi.org/10.1109/LGRS.2023.3272353
[31] Kaur G, Afaq Y. Developments in deep learning for change detection in remote sensing: A review. Transactions in GIS. 2024 Apr;28(2):223-57. https://doi.org/10.1111/tgis.13133
[32] Cheng G, Huang Y, Li X, Lyu S, Xu Z, Zhao H, Zhao Q, Xiang S. Change detection methods for remote sensing in the last decade: A comprehensive review. Remote Sensing. 2024 Jun 27;16(13):2355. https://doi.org/10.3390/rs16132355

[33] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In2010 20th international conference on pattern recognition 2010 Aug 23 (pp. 3121-3124). IEEE. https://doi.org/10.1109/ICPR.2010.764