Remote Sensing Applications: Society and Environment 39 (2025) 101698

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

High precision building detection from satellite imagery with a novel SDBN-HCWO method

Md Helal Miah [a,b,*], Shuanggen Jin [c,d], Mayin Uddin Jubaid [b], Md Altab Hossin [e]

[a] School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou City, Henan, China
[b] School of Aerospace Engineering, Shenyang Aerospace University, Shenyang, Liaoning, China
[c] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China
[d] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, 454003, China
[e] School of Innovation and Entrepreneurship, Chengdu University, Chengdu, Sichuan, China

ARTICLE INFO

Keywords:
Building detection
Satellite imagery
Edge identification
Multi-layer optimization
Image quality

ABSTRACT

Detecting buildings from satellite imagery presents challenges related to computational efficiency, model adaptation, and occlusion. This paper introduces a novel method called the Secant Deep Belief Network-Hyperbolic Cosine Whale Optimization (SDBN-HCWO) for building detection in satellite images. The research utilizes the SDBN-HCWO method to enhance building detection accuracy in satellite images. It addresses challenges like computational efficiency, occlusion, and dataset adaptation. The method integrates a multi-layer structure, including Hyperbolic Cosine Prey Encircling for edge identification, Shrinking Encircle for optimal edge linking, and Secant Object Detection for accurate identification. Additionally, a Densely Connected Convolutional Network (DCCN) and Depth-wise Separable Convolution (DSC) optimize feature extraction, reducing computational costs. The model is evaluated on both quantitative and qualitative metrics, ensuring high accuracy and low false positive rates. The research findings demonstrate that the SDBN-HCWO method significantly improves building detection accuracy in satellite imagery. It enhances detection efficiency by integrating Discrete Latent Deep Reinforcement Learning and a bubble-net mechanism, reducing false positives by 58 %. The model outperforms conventional approaches, achieving an 18 % increase in PSNR, 34 % rise in CA, and 19 % reduction in training time. High AP scores (90.40 %–92.67 %) confirm its reliability, though challenges persist in medium-damage detection. It surpasses YOLOv3, YOLOv4, and Faster R-CNN in accuracy and efficiency. This research significantly advances building detection in satellite imagery, facilitating more accurate urban planning, disaster response, and environmental monitoring.

1. Introduction

Detecting buildings from aerial imagery is a complex task due to various challenges. These challenges include obstructions from nearby trees, shadows cast by surrounding structures, a variety of rooftop textures and colours, and the wide range of building shapes and sizes. These factors often make it difficult for modern models to accurately identify and outline building boundaries. Access to high-resolution aerial imagery datasets is crucial for evaluating different detection techniques and has fuelled growing interest in the

* Corresponding author.
E-mail address: helal.sau.12030704@gmail.com (M.H. Miah).

https://doi.org/10.1016/j.rsase.2025.101698
Received 19 April 2025; Received in revised form 20 July 2025; Accepted 23 August 2025
Available online 25 August 2025

applications of aerial imagery within machine learning and computer vision research. This technology plays a crucial role in geographic information systems (GIS), which are vital for urban planning, enabling city planners to visualize and manage urban growth effectively. Furthermore, it aids in disaster monitoring by providing timely and accurate data about structures in affected areas, facilitating better response strategies and recovery efforts. As the demand for sophisticated mapping and analysis tools continues to grow, building detection remains at the forefront of innovation in remote sensing and spatial analysis. (Zhang, 1999). The accurate and efficient identification of buildings from satellite images is essential, but it poses significant challenges, such as reducing distortions and minimizing noise while ensuring high precision (Zeng et al., 2022). Deep learning (DL) has become a powerful tool, providing structured methodologies for knowledge-based decision-making. By extracting patterns from large datasets, DL models can make accurate predictions for unseen inputs, which makes them highly effective for tasks such as object detection (Ayemowa et al., 2024). One of the main challenges in deep learning-based building detection is the selection and optimization of model parameters, which directly affects performance. To tackle this issue, researchers have increasingly adopted bio-inspired optimization techniques. By integrating nature-inspired algorithms, they aim to enhance parameter selection and improve the accuracy of the models (Wu et al., 2024; Luo et al., 2021). These methods have shown great promise in enhancing deep learning models, making them more robust and efficient for detecting buildings in satellite images.

Building detection accuracy has been a significant area of research in remote sensing and computational intelligence, with numerous techniques developed to improve both precision and efficiency. Huang et al. (2025) developed PP-BCD, a deep learning-based building change detection framework utilizing pixel-wise and patch-wise fusion strategies. Tested on Pleiades imagery, PP-BCD showed improved accuracy, outperforming traditional methods and reducing sample collection costs (Huang et al., 2025). Luis et al. developed a YOLOv8-based deep learning model for building damage detection using high-resolution satellite imagery. Their approach, optimized with SGD, improved prediction reliability and speed, offering a scalable tool for real-time disaster assessment (Luis et al., 2025). Bhardwaj et al. introduced an integrated ResNet-U-Net model for post-disaster building damage assessment using satellite imagery. The model achieved 96.91 % accuracy on unbalanced datasets, improving efficiency and reliability for disaster response applications (Bhardwaj et al., 2025). Khankeshizadeh et al. proposed the WETUM model for post-earthquake building damage assessment using UAV data. The model achieved a 78.26 % damage detection rate, outperforming deep learning and machine learning models by integrating spectral and geometrical features (Khankeshizadeh et al., 2024). Xie et al. (2024) developed a landslide extraction framework using a two-branch network and context association features. Their method improved accuracy, achieving a 0.92 %–16.94 % higher IoU compared to 17 contemporary deep learning methods (Xie et al., 2024). Sun et al. developed a deep convolutional network model for detecting dumpsites from high-resolution satellite imagery. Their approach improved detection efficiency, reducing investigation time by 96.8 %, enabling large-scale analysis of global waste distribution factors (Sun et al., 2023). Chen et al. proposed a novel approach integrating aerial imagery with BIM for automated concrete defect detection. Their method, using bundle registration, improved accuracy by reducing false positives and enhancing 3D defect reconstruction, achieving 6.4 % higher IoU (Chen et al., 2023a). Singh and Nongmeikapam employed Deep-Unet for semantic segmentation of satellite images, enhancing land cover mapping. Their method, integrating FAAGKFCM and SLIC Superpixel, improved accuracy, achieving 90.6 % global accuracy and outperforming existing models (Singh and Nongmeikapam, 2022). Tanim et al. developed a machine-learning framework for urban flood detection using Sentinel-1 satellite imagery and ground reports. Their unsupervised model improved accuracy (0.87) over RF, SVM, and MLC, enhancing rapid flood mapping and risk management (Tanim et al., 2022). PushpaRani et al. proposed a deep learning approach using U-Net for geological information extraction from satellite imagery. Their method, enhanced by data augmentation, improved accuracy, achieving a 92 % F1 score and 95 % precision, reducing manual processing time (PushpaRani et al., 2024). Sirko et al. developed a U-Net-based pipeline to improve building detection accuracy from high-resolution satellite imagery. Validated on 100k African images, their approach used mix-up augmentation and soft KL loss, enhancing segmentation and contributing to the Open Buildings dataset (Sirko et al., 2021). Corbane et al. developed a deep-learning framework using convolutional neural networks for global human settlement mapping from Sentinel-2 imagery. Their model, validated on 277 sites, improved accuracy in built-up area extraction, enhancing urbanization and sustainability assessments (Corbane et al., 2021). Abdollahi et al. introduced the Seg-Unet model, combining Segnet and Unet for building segmentation from high-resolution aerial images. Their method improved accuracy to 92.73 % on the Massachusetts dataset, outperforming FCN, Segnet, and Unet in precise building extraction (Abdollahi et al., 2022). Rahnemoonfar et al. introduced FloodNet, a high-resolution UAV imagery dataset aimed at post-flood scene understanding through semantic segmentation and visual question answering. While primarily developed for disaster assessment, its high-resolution data and annotated flooded buildings make it highly relevant for building detection tasks. The dataset supports training deep learning models to recognize and segment partially submerged or occluded structures, a challenge commonly encountered in building detection, especially in complex or post-disaster urban environments (Rahnemoonfar et al., 2021).

Sharma and Singhai developed a Modified GrabCut Partitioning algorithm for building detection, incorporating bio-inspired optimization to improve segmentation accuracy. Their method demonstrated significant enhancements in energy minimization, which typically struggles in complex boundary scenarios due to its dependence on local optimization. However, the study lacks evaluation across diverse urban datasets and varying image conditions, indicating the need for broader generalizability testing (Sharma and Singhai, 2021). Additionally, Argyrou and Agapiou explored a computational intelligence model for analysing remote sensing images, categorizing methodologies into assessment algorithms and neural networks. The study also reviewed expansion rankings and future advancements in remote sensing image registration, offering insights into the evolving perspectives in this field (Argyrou and Agapiou, 2022). A novel building detection technique was proposed by Huang et al. In their study, vegetation regions were first identified from satellite images. They used local directional fuzzy landscapes to analyze shadow regions and illumination direction in order to determine the presence of buildings. Subsequently, a graph-based algorithm was employed to differentiate between foreground and background pixels across the entire landscape, resulting in improved detection accuracy (Huang et al., 2020). A

bio-inspired method known as the Jellyfish Search Algorithm (JSA) was explored by Chou and Molla as an optimization technique for building detection. Due to the vast amount of remote sensing data, JSA was recognized as a robust metaheuristic method, drawing inspiration from the food-finding behavior of jellyfish in the ocean. The study emphasized the algorithm's effectiveness, demonstrating that it outperformed several established metaheuristic approaches across various benchmark functions and real-world applications (Chou and Molla, 2022). The Manta Ray Foraging Optimization (MRFO) technique was used by Rai et al. in conjunction with the Radial Basis Function Neural Network (RBFNN) to efficiently extract hidden building information. The trained neural network was essential in identifying concealed data, and the use of spectral indices further improved its performance (Rai et al., 2023). Furthermore, research has investigated the use of super-resolution techniques for large-scale building detection. Chen et al. utilized both low- and high-resolution open-source datasets for model segmentation, applying color normalization and a super-resolution algorithm to enhance image quality and improve instance segmentation accuracy. They developed a Mask R-CNN model based on the MPViT backbone, which demonstrated better object-wise performance compared to mainstream instance segmentation approaches. However, the study identified challenges in applying super-resolution models across different magnification scales and in improving the separation of closely located buildings. Additionally, it was observed that unsupervised learning-based building extraction did not significantly improve the effectiveness of the proposed method (Chen et al., 2023b).

Günen, Öztürk, and Aliyazıcıoğlu proposed a modified weighted differential evolution-based viewshed analysis (mWDE-WS) to optimize the visibility of historical structures in Türkiye's Kromni Valley. Their method improved visibility from 64 % to 84.37 %. However, the model's adaptability to other regions and exclusion of cultural or symbolic factors limits broader generalization (Günen et al., 2025). Günen proposed a novel framework combining VHR imagery, topographic indices, and machine learning for accurate rooftop detection. Random forest achieved the highest performance (98.72 % accuracy) using full feature sets. The study's key limitation lies in reduced model effectiveness for RGB-only data and potential generalization issues across varying urban landscapes or less-structured environments (Günen, 2024).

Overall, Existing research has introduced various bio-inspired and deep learning techniques aimed at improving building detection accuracy. However, challenges still exist in optimizing computational efficiency, adapting models to diverse datasets, and addressing occlusion-related issues in urban environments. The accuracy for building detection using various existing deep learning methods is generally 80 % or lower. Future advancements should focus on integrating multi-scale feature extraction and hybrid optimization frameworks to further enhance detection precision. One promising approach is the Secant Deep Belief Network-based Hyperbolic Cosine Whale Optimization (SDBN-HCWO), which aims to improve building detection by addressing local optima problems and achieving a balanced trade-off between exploration and exploitation. The SDBN-HCWO method combines bio-inspired Hyperbolic Cosine Whale Optimization with the Secant Deep Belief Network (SDBN), which consists of both visible and hidden layers. Specifically, it employs three hidden layers, each serving a different purpose in refining object detection.

- The first hidden layer uses a Hyperbolic Cosine Prey Encircling-based Best Edge Identification model to pinpoint the most relevant edges, laying a strong foundation for precise detection.
- The second hidden layer applies a Shrinking Encircle and Spiral Update-based Optimal Edge Linking model to effectively connect the detected edges while reducing distortions and false detections.
- Finally, the third hidden layer utilizes the Secant Object Detection model to achieve accurate and reliable building identification.

The SDBN-HCWO method evaluates both quantitative and qualitative performance based on the best fitness values, ultimately ensuring higher accuracy and a lower false positive rate in building detection. This comprehensive approach significantly enhances detection performance, making it a promising solution for object identification in satellite images. In addition, the SDBN-HCWO model innovatively integrates the Densely Connected Convolutional Network (DCCN) for enhanced feature extraction and computational efficiency, departing from traditional deep learning structures. By linking every layer to all preceding layers, DCCN improves feature reuse and gradient flow, addressing vanishing gradients and feature loss. The model uses dense blocks and transition layers with batch normalization and pooling to optimize feature extraction while maintaining detection accuracy, allowing efficient processing of large-scale input data. Additionally, Depth-wise Separable Convolution (DSC) reduces computational costs and parameters, enabling real-time object detection without sacrificing accuracy. Finally, the H-Swish activation function improves gradient propagation and convergence speed, enhancing reliability and adaptability for real-time applications in resource-constrained environments.

The research innovation are followings.

- Development of the SDBN-HCWO Framework: Introduced a novel hybrid architecture combining Secant Deep Belief Network (SDBN) with Hyperbolic Cosine Whale Optimization (HCWO) to enhance building detection accuracy and computational efficiency.
- Integration of Densely Connected Convolutional Network (DCCN): Leveraged dense connections between layers to enhance feature reuse, gradient flow, and reduce the impact of vanishing gradients and feature loss, thereby boosting learning efficiency and accuracy.
- Use of Depth-wise Separable Convolution (DSC): Applied DSC to significantly reduce computational cost and model parameters, enabling real-time object detection without compromising precision.
- Implementation of H-Swish Activation Function: Adopted the H-Swish activation to improve gradient propagation and convergence speed, enhancing the model's adaptability and reliability for real-time applications in resource-limited environments.

- Performance Gains Over Existing Methods: Demonstrated superior accuracy and lower false positive rates compared to traditional deep learning methods, with the proposed model achieving building detection performance above 80 %, addressing challenges like occlusion, diverse datasets, and local optima.
- Balanced Exploration-Exploitation Strategy: The hybrid HCWO mechanism ensures a balanced trade-off between exploration and exploitation, avoiding convergence to local optima and improving the global search capabilities of the detection model.

2. Methods and models

2.1. SDBN-HCWO model

The rapid advancements in satellite communication have made high-resolution satellite imagery increasingly important across various fields. However, detecting buildings from single monocular images is a complex and computationally demanding task because these images do not provide clear information about a building's size, shape, or color. Traditional methods for building detection rely on statistical relationships between training and test image data, which can limit their effectiveness in accurately identifying structures. To overcome these limitations, Mayank Dixit and colleagues introduced the Dilated-ResUnet deep learning model. This model incorporates dilated convolution and residual blocks to improve feature extraction from satellite imagery (Dixit et al., 2021). While this approach improved detection accuracy, it did not effectively reduce false detections. In a related study, Lingling Fang, Xiyue Fang, and Liang proposed an optimization model that utilizes a nonlinear binary grasshopper whale algorithm. This model incorporates a novel position updating strategy to achieve a higher accuracy rate (Fang and Liang, 2023a). Although this approach improved detection accuracy, it did not sufficiently reduce false detections. In a similar vein, Lingling Fang and Xiyue Fang proposed an optimization model that utilizes a nonlinear binary grasshopper whale algorithm. They incorporated a novel position updating strategy to achieve a higher accuracy rate (Fang and Liang, 2023b).

The Secant Deep Belief Network (SDBN) is a cutting-edge machine learning approach designed for high-precision building detection from satellite imagery, particularly in urban environments where structural complexity and occlusions challenge traditional detection methods. By integrating the deep learning principles of Deep Belief Networks (DBNs) with the Secant method's optimization capabilities, SDBN significantly enhances feature extraction and convergence efficiency. The model processes high-resolution satellite images to capture intricate spatial patterns and architectural variations, learning hierarchical representations of building features with minimal manual intervention. Compared to conventional convolutional networks, SDBN achieves superior detection accuracy, reduced false positives, and faster training convergence due to its hybrid architecture. Its applications extend to urban planning, disaster response, and smart city development, where accurate and automated building footprint extraction is critical (Mozafari and Moattar, 2025).

The Hyperbolic Cosine Whale Optimization (HCWO) Algorithm is an advanced metaheuristic optimization technique tailored for high-precision building detection from satellite imagery. This method combines the natural foraging behavior of humpback whales, as modelled in the traditional Whale Optimization Algorithm (WOA), with the hyperbolic cosine function to improve exploration and exploitation balance during optimization. When applied to satellite image analysis, HCWO enhances the tuning of deep learning model parameters, enabling more accurate extraction of building footprints in complex urban landscapes. Its non-linear search dynamics help avoid local minima, allowing the algorithm to optimize feature selection and classifier weights more effectively. By integrating HCWO with deep neural networks, the approach achieves superior segmentation accuracy, especially in densely built or visually noisy regions. This makes HCWO highly valuable in geospatial intelligence, urban monitoring, and post-disaster infrastructure assessment (Xu and Zhang, 2024).

To achieve an optimal balance between exploration and exploitation while avoiding local optima, this research developed a novel method called SDBN-HCWO for building detection from satellite images. This approach features an innovative mechanism that controls the positioning of whales by modifying the Hyperbolic Cosine function during the optimization process. Additionally, the

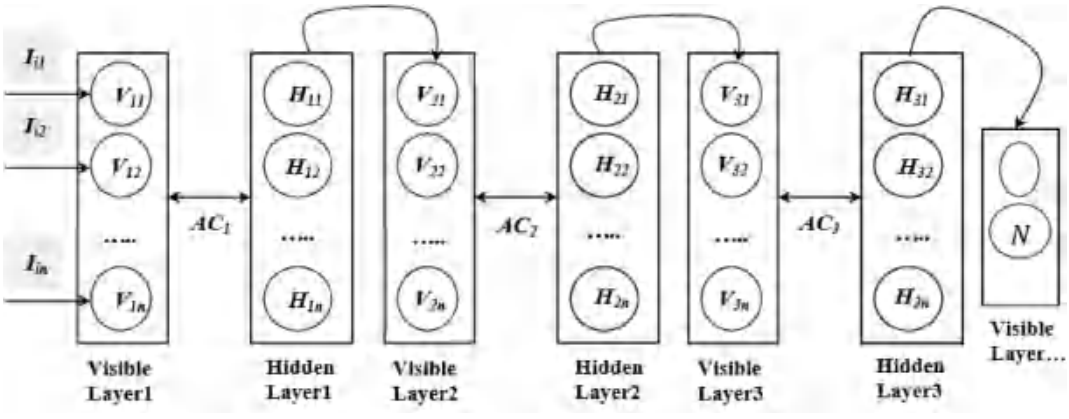

Fig. 1. Structure of SDBN-HCWO method.

method incorporates rectification features to fine-tune the adjustments of search agents within the search space. It leverages the Secant Deep Belief Network to enhance precision. These improvements ensure a well-balanced trade-off between exploration and exploitation, allowing the model to effectively identify buildings while maintaining curvature constraints. By refining the search process, the SDBN-HCWO method facilitates increased accuracy and robustness in detecting structures from satellite imagery. This represents a significant advancement in deep learning-based object detection. The structural framework of the SDBN-HCWO model is illustrated in Fig. 1, which depicts its systematic approach to optimizing building detection.

Fig. 1 illustrates the structural framework of the SDBN-HCWO method, which consists of hidden states and a visible state, featuring inter-layer connections while avoiding intra-layer links. The visible state, also known as the visible layer, represents the input layer where observed data is introduced. In contrast, the hidden states, or hidden layers, carry out the core processing tasks. Specifically, the SDBN-HCWO model comprises three hidden layers and two visible layers, providing a structured approach to building detection. Initially, satellite images are input into the visible layer, where a bio-inspired optimization process is performed using Hyperbolic Cosine Whale Optimization across the three hidden layers. In the first hidden layer, the best edge identification occurs during the prey-encircling phase, where the "prey" represents the targeted buildings. The identified edges are then sent back to the visible layer, serving as input for the second hidden layer, where the exploitation phase optimally links the detected edges while minimizing false positives. In the final hidden layer, the exploration phase refines object detection to ensure high accuracy. Ultimately, based on the proximity of the detected objects to the current estimates, the secant equation is applied in the visible layer to finalize the building detection process. This structured, multi-layered approach significantly enhances detection accuracy while reducing computational errors.

Table 1 presents a comparative overview of the control parameters utilized by the proposed SDBN-HCWO model in contrast with several existing optimization algorithms, highlighting the superior configurability and adaptability of the proposed approach. Unlike CSA-PS and FOA-PS, which are parameter-free and thus less flexible for fine-tuning, TLBO-PS introduces minimal control through a fixed teaching factor, and PSO-PS relies on standard acceleration coefficients ($c1 = c2 = 1.80$) and inertia weight ($\omega = 0.60$). JADE-PS incorporates adaptive parameters such as crossover rate ($c \in$ (Luo et al., 2021; Sharma and Singhai, 2021)) and selection probability ($p \in$ [5 %N, 20 %N]). In contrast, the SDBN-HCWO framework integrates a comprehensive set of tunable parameters, including population size (30–50), maximum iterations (100–200), convergence coefficient (ranging from 2 to 0), and a hyperbolic cosine scaling factor (1.5–3.0), which collectively enhance the model's optimization performance. Furthermore, the model supports fine-grained control of deep learning parameters, such as learning rate (0.001–0.01), number of layers (3–5), hidden units per layer ([500, 300, 100]), batch size (32–128), epochs (50–200), and a secant gradient threshold (10^{-6}). This rich parameterization enables the SDBN-HCWO model to effectively adapt to the complexities of high-precision building detection from satellite imagery, offering improved accuracy and robustness over traditional methods.

2.2. Problem formulation

The optimization process in the SDBN-HCWO method starts with initializing training vectors in the visible units, which correspond to the visible layer. During this stage, sample images used for simulation are placed in the visible layer, providing the foundational input for the optimization procedure. Each whale position within the population represents a potential solution and is encoded as a position vector made up of pixel intensity values. These values, which range from 0 to t, establish the initial thresholds for processing the satellite images. This structured initialization ensures that the optimization process can effectively explore and exploit image features, enabling precise and efficient building detection. By representing each image as a set of pixel intensity variations, equation (1) creates a robust framework that enhances detection accuracy while maintaining computational efficiency.

$$O_i = (I_{i1}, I_{i2}, \ldots\ldots, I_{in}) \tag{1}$$

Table 1
A comparison of the control parameters of SDBN-HCWO with existing algorithms.

Methods	Control Parameters
CSA-PS (Günen et al., 2024)	No parameter
TLBO-PS (Günen et al., 2024)	Teaching factor = (Chen et al., 2023a)
PSO-PS (Günen et al., 2024)	$c1 = c2 = 1.80$, $\omega = 0.60$
FOA-PS (Günen et al., 2024)	No parameter
JADE-PS (Günen et al., 2024)	$c \in$ (Luo et al., 2021; Sharma and Singhai, 2021), $p \in$ [5 %N,20 %N]
SDBN-HCWO	Population size (Fang and Liang, 2023b; Mozafari and Moattar, 2025; Xu and Zhang, 2024; Günen et al., 2024; Cao et al., 2023; Han et al., 2022; Hossain et al., 2021; Li et al., 2024; Langerman et al., 2020; Ju et al., 2024; Shi et al., 2021; Senussi and Kang, 2024; Massachusetts Buildings Dataset; Deep learning for satellite image; Kononchuk et al., 2022; Basha and Logu, 2024; Dhanaraj et al., 2021; Balyan et al., 2022; Javed et al., 2021; Uddin et al., 2022; Sanhudo et al., 2021), Maximum iterations [100–200], Convergence coefficient [2 to 0], Hyperbolic cosine scaling factor [1.5–3.0], Secant gradient threshold [10^{-6}], Learning rate [0.001–0.01], Number of layers (Ayemowa et al., 2024; Wu et al., 2024; Luo et al., 2021), Hidden units per layer [500, 300, 100], Batch size [32–128], Epochs [50–200].

where the parameter '*n' represents the population size of whales, which corresponds to the total number of sample images used for simulation. Likewise, 'I_{i1}*' refers to the actual image under consideration in the detection process. To ensure a diverse and well-distributed search space, the initial positions of the whales (symbolizing the sample images) are generated randomly using a predefined mathematical formulation. This randomized initialization prevents premature convergence and enhances the model's ability to explore multiple potential solutions effectively. By incorporating this strategy, the optimization process gains a strong foundation for refining the detection of buildings in satellite images, ensuring improved accuracy and robustness in identifying structural patterns. The subsequent equation provides the computational framework for this initialization, laying the groundwork for further refinement in the detection process.

$$\mathrm{I_{ij}} = \mathrm{Res_{min}} + (\mathrm{Res_{max}} - \mathrm{Res_{min}}) \times random(0,1), I_{ij} \in 0_i \& j = 1, 2, \ldots..m \tag{2}$$

where 'Res_{min}' and 'Res_{max}' denote the minimum and maximum intensity levels of the sample images in the histogram, respectively. These values establish the range of pixel intensities crucial for image processing in the SDBN-HCWO framework. The model incorporates three hidden layers, each assigned a specific function to enhance building detection accuracy. In the first hidden layer, the prey encircling phase is executed, where the system identifies and localizes potential buildings. The second hidden layer then carries out the exploitation phase, refining edge detection and minimizing false positives. Finally, the exploration phase in the third hidden layer ensures optimal object identification by adjusting the model's focus on significant features. As the process transitions between hidden layers, updates are systematically applied to the corresponding hidden or visible layers, following a structured update mechanism outlined in the subsequent equations. This multi-layered approach enhances detection precision while maintaining computational efficiency.

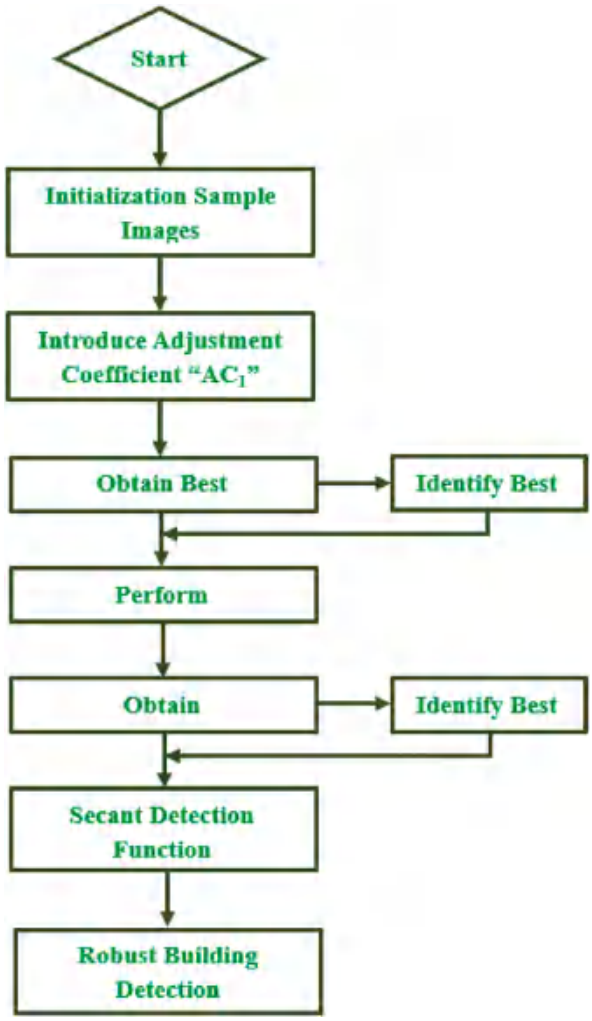

Fig. 2. Flow chart of SDBN-HCWO.

$$\text{Prob}(Hid_i = 1|Vis) = \sigma\left(b_j + \sum_{i}^{1} Vis_i W_{ij}\right) \tag{3}$$

$$\text{Prob}(Vis_i = 1|Hid) = \sigma\left(a_i + \sum_{j}^{1} Hid_i W_{ij}\right) \tag{4}$$

From Equations (3) and (4), the update mechanism between the visible and hidden layers in the SDBN-HCWO method is governed by the sigmoid function outcomes. Given the visible layer '*Vis*', the corresponding hidden layer 'Hid_i' is updated based on the sigmoid function result 'σ' and the bias term 'b_j' associated with 'Hid_i'. Similarly, when the hidden layer '*Hid*' is provided, the visible layer 'Vis_i' is updated using the sigmoid function outcome 'σ' along with the bias parameter 'a_i' of 'Hid_j'. These iterative updates ensure effective feature learning and enhance the accuracy of building detection. As outlined in the SDBN-HCWO framework, a structured three-step process is executed within the hidden layers: the first layer performs the prey encircling phase, the second layer handles the exploitation phase, and the third layer conducts the exploration phase. This hierarchical processing approach optimizes the balance between detection precision and computational efficiency. The complete flow of the SDBN-HCWO method is visually represented in Fig. 2, illustrating the sequential interactions between layers and the systematic refinement of building detection outcomes.

As illustrated in Fig. 2, the prey encircling phase is crucial for determining the optimal position vector of the prey, which in this context refers to the best-identified building images. This phase lays the groundwork for the subsequent stages of processing by identifying the most relevant structural features. Following this, two key procedures are implemented: the exploitation phase and the exploration phase. The exploitation phase focuses on connecting the best edges to enhance structural continuity, while the exploration phase is responsible for searching and refining the detected buildings. These two phases work in tandem to achieve a finely tuned building detection outcome. By systematically adjusting the positions of the detected structures, the model ensures greater accuracy and minimizes false positives. A more detailed explanation of the processes within the three hidden layers, including their specific roles in enhancing detection efficiency, is provided in the following subsections.

2.3. Developed best edge identification model

In the first hidden layer, the prey encircling phase is performed, during which the position of the prey is determined based on its proximity to the target. To improve this process, Di Cao et al. introduced an enhanced dynamic opposite learning approach along with an adaptive encircling prey mechanism. They utilized a novel whale optimization technique to facilitate classification based on mechanical building identification characteristics (Cao et al., 2023). Despite advancements in the field, the accuracy of classification has not improved. To address this limitation, the SDBN-HCWO method introduces Hyperbolic Cosine Prey Encircling for edge identification. In this phase, each humpback whale is represented by sample images, and the movement of the fine agent, which denotes edges, is tracked through the remaining sample images. However, if the positional adjustments during the search are not sufficient, the entire sample image may fail to undergo the necessary behavioural changes needed to identify the optimal building edges. To overcome this challenge, the method employs dual adjustment coefficients, AC_1 and AC_2, which dynamically regulate changes in the coefficient values, thereby enhancing edge detection accuracy. The mathematical formulation (Equations (5) and (6)) of this behavior further refines the edge identification process, ensuring more precise classification outcomes.

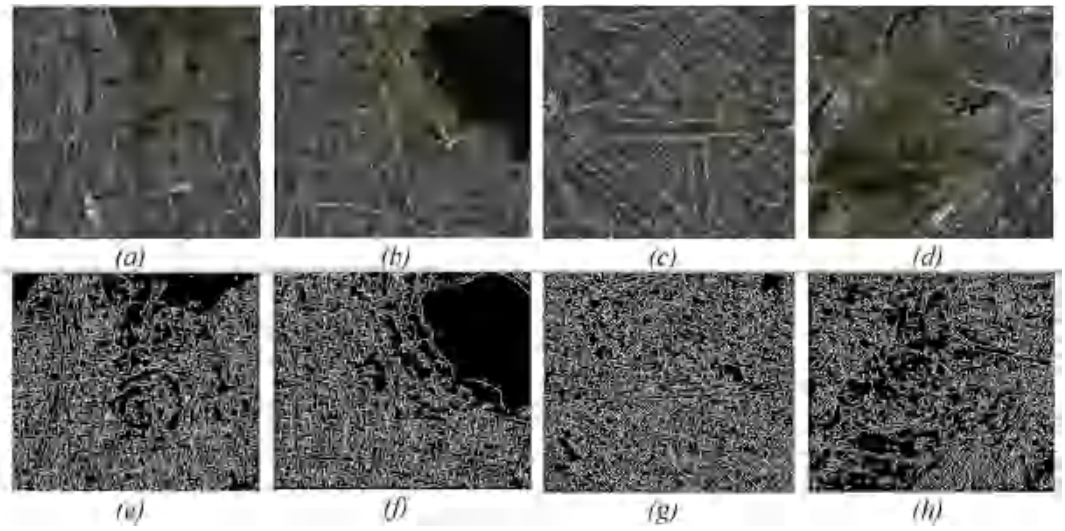

Fig. 3. (a), (b), (c), (d) – sample input images (e), (f), (g), (h) – Output result after applying Hyperbolic Cosine Prey Encircling-based Best Edge Identification.

$$PV_i(t+1) = \frac{PV_{best}(t) - CV_1 D_i}{AC_1} \tag{5}$$

$$D_i = \frac{CV_2 PV_{best}(t) - PV_i(t)}{AC_1} \tag{6}$$

Equations (5) and (6) describe the relationship between key variables in the optimization process. Specifically, $PV_{best}(t)$ represents the best position vector identified in the current iteration t, while $PV_i(t)$ denotes the position vector of the individual whale (i.e., sample image) at that same iteration. The coefficient vectors CV_1 and CV_2 are used to guide the search process, influencing the movement and adjustment of the whales. Additionally, D_i captures the difference between the best position vector $PV_{best}(t)$ achieved up to that point and the current position vector $PV_i(t)$ for the ith whale at iteration t. This variation helps track the progress and adjustments made during the optimization. The best edges are determined by analysing the changes in these position vectors, with the coefficient vectors CV_1 and CV_2 being mathematically defined as outlined below.

$$CV_1 = (2av - 1)dv = (2av - 1)(2 - t)\cosh\left(\frac{2\pi}{\max(t)}\right) \tag{7}$$

$$CV_2 = 2 \times av \tag{8}$$

From the above equations (7) and (8), on the basis of arbitrary vector av and decreasing vector dv, the coefficient vectors CV_1 and CV_2 are obtained. To improve the positioning of search agents (i.e., edges) and avoid local optima caused by insufficient knowledge of the best possible search agent, this research employed a Hyperbolic Cosine function in the decreasing vector (dv) to guide the optimization process. This function ensures a smooth balance between exploration and exploitation throughout the overall iteration process, thereby enhancing the Peak Signal-to-Noise Ratio (PSNR). Fig. 3 illustrates the results of the best edge identification using the Hyperbolic Cosine Prey Encircling-based Best Edge Identification model.

Fig. 3 shows the results of best edge identification using the Hyperbolic Cosine Prey Encircling model. The figure includes four sample images, labelled (a) to (d), each representing different input scenarios. The corresponding edge identification results for these images are displayed in panels (e) to (h). This highlights the model's effectiveness in accurately detecting edges across various samples. This visual representation emphasizes the model's ability to optimize edge detection and produce reliable outputs in diverse contexts.

2.4. Developed optimal edge linking model

In the exploitation phase, after identifying the best edges during the prey encircling phase in the first hidden layer, the focus shifts to optimal edge linking. This phase is executed using the bubble-net mechanism, which involves two key tasks: the shrinking encircling task and the spiral updating task. Both tasks are performed in the second hidden layer. In this context, whales symbolize the search agents that utilize the bubble-net mechanism to encircle the prey while forming bubbles in a helix-shaped pattern. In the SDBN-HCWO model, the edges that are detected are linked together to minimize errors and ensure more accurate edge identification. The bubble-net mechanism is vital as it creates distinctive helix-shaped bubbles that connect the relevant edges. Additionally, the imaging coefficient vectors CV_1 and CV_2 are adjusted in a random manner within the shrinking encircling technique, with shifts occurring from 2 to 0 to fine-tune the coefficients. This dynamic shifting mechanism facilitates smooth horizontal motion across all iterations. Following this,

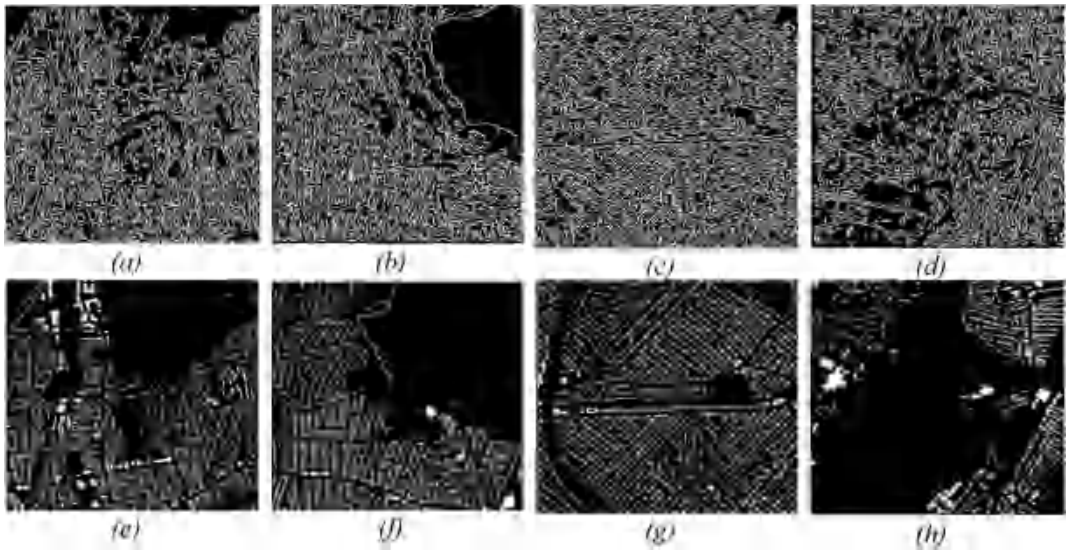

Fig. 4. (a), (b), (c), (d) – edge detected images (e), (f), (g), (h) – Output result after applying Shrinking Encircle and Spiral Update model.

the spiral updating equation is applied during the exploitation phase. This involves using an adjustment coefficient to enhance the edge sequence, resulting in a helix-like structure that improves edge continuity and consistency throughout the process.

$$PV_i(t+1) = \frac{\left[D_i' e^{bl} \cosh(2\pi l) + PV_{best}(t)\right]}{AC_2} \tag{9}$$

$$D_i' = PV_{best}(t) - PV(t) \tag{10}$$

Equations (9) and (10) describe the distance D_i', which represents the difference between the best position vector $PV_{best}(t)$ at iteration t, with b and l denoting the parameters for a logarithmic spiral shape and a random number within the range of [−1, 1], respectively. The logarithmic spiral is used to connect the best or relevant edges while effectively removing isolated ones, which minimizes errors in detecting objects during the final stages of the process. This technique ensures more accurate and consistent edge identification. Fig. 4 visually presents the results of optimal edge linking, achieved through the Shrinking Encircle and Spiral Update model. This demonstrates the effectiveness of the method in refining edge connectivity and enhancing the overall detection process.

Fig. 4 (a)–(d) show the input images after edge detection, while the results produced by the Shrinking Encircle and Spiral Update model are illustrated in Fig. 4(e)–(h).

2.5. Developed SDBN-HCWO object detection model

The proposed SDBN-HCWO method executes a prey-searching phase that corresponds to object detection, enabling accurate identification of objects. Han et al. introduced a Mask Region-CNN for building detection; however, their model did not consider the time needed for weight updates during classification, which ultimately impacted the overall accuracy (Han et al., 2022). To overcome this limitation, the Secant Object Detection Model has been developed. The main goal of the exploration phase is to thoroughly investigate the entire search space (i.e., the sample images for analysis) beyond the current best solution in order to identify a more optimal option in a way that is computationally efficient. This process is facilitated by updating the position of coefficient vectors (CV_1, CV_2), as outlined below.

$$PV_i(t+1) = \frac{PV_{rand} - CV_1 D_i}{AC_2} \tag{11}$$

$$D_i = \frac{CV_2 . PV_{rand} - PV_i(t)}{AC_1} \tag{12}$$

From equations (11) and (12), 'PV_{rand}' represents a randomly selected position vector for the current iteration (t). Based on the position of the coefficient vectors (CV_1, CV_2), the randomly chosen position vector for the corresponding image is updated iteratively. This process continues until the termination condition is met. To find a solution to the secant equation, the approach takes into account how close the target is to the current estimate while ensuring that the curvature condition in the visible layer is satisfied, as outlined below.

$$\mathrm{Res}_{i+1} = \left(I - \gamma_i PV_i \mathrm{Sec}_i^T\right) \mathrm{Res}_i \left(I - \gamma_i \, \mathrm{Sec}_i PV_i^T\right) + \gamma_i PV_i PV_i^T \tag{13}$$

$$PV_i = PV_{best} - PV, \gamma_i = \frac{1}{PV_i^T \, \mathrm{Sec}_i} \tag{14}$$

According to equations (13) and (14), a building is identified when the resultant class value exceeds zero; otherwise, no building is detected.

The algorithm for the SDBN-HCWO is outlined below.

Algorithm I.

Input: Dataset '*DS*', Images '$I = I_1, I_2, \ldots, I_n$'

Output: Robust building detection

1: **Initialize** '*n*', arbitrary vector '$av = [0,1]$', decreasing vector '$dv = [2 \rightarrow 0]$', iteration '*t*'
2: **Initialize** adjustment coefficients '$AC_1 = 2.5$', '$AC_2 = 1.5$'
3: **Begin**
4: **For** each Dataset '*DS*' with Images '*I*'
//visible layer
5: Formulate series of pixel intensity values as given in (1)
6: Generate initial position of the whales (i.e., sample images considered for simulation) as given in (2)
7: Formulate switching between hidden layers or visible layer as given in (3) and (4)
//Prey encircling phase (i.e., building edge detection) – Hidden layer 1
8: Observe the behavior of the whales (i.e., sample images considered for simulation) as given in (5) and (6)
9: Evaluate the coefficient vectors 'CV_1' and 'CV_2' as given in (7) and (8)
10: **Return best edges**

(*continued on next page*)

(*continued*)

Input: Dataset '*DS*', Images '$I = I_1, I_2, \ldots, I_n$'

Output: Robust building detection

//Exploitation phase (i.e., edge linking) – Hidden layer 2
//shrinking encircling technique
11: Fine-tune coefficient vectors 'CV_1' and 'CV_2' arbitrarily from '2' to '0'
//spiral updating techniques
12: Perform spiral updating employing adjustment coefficient as given in (9) and (10)
13: **Return optimal edge linkages**
//Exploration phase (i.e., object detection) – Hidden layer 3
14: Explore overall search space (i.e., sample image for analysis) to obtain the current best solution as given in (11) and (12)
15: **Return predicted results**
//Visible layer (i.e., building detection)
16: Evaluate resultant value as given in (13) and (14)
17: **If** '$Res_{i+1} > 0$' then '*FRes* = building *detected*'
18: **Else** '*FRes* = no building *detected*'
19: **End if**
20: **End for 21**: **End**

Algorithm I outlines the process of building detection while minimizing the false positive rate using the Secant Deep Belief Network with Hyperbolic Cosine Whale Optimization (SDBN-HCWO) method, which incorporates bio-inspired optimization within a deep learning framework. This approach employs a novel Secant Deep Belief Network (SDBN) consisting of a visible layer and three hidden layers, each performing specific operations related to the Hyperbolic Cosine Whale Optimization technique. The visible layer processes multiple input images and sends them to the first hidden layer, where edge detection is conducted using the Hyperbolic Cosine Prey Encircling function. The optimal edges identified are then refined in the second hidden layer, which establishes relationships between the edges of the corresponding images. This is achieved by fine-tuning the initialized coefficient vectors through the Shrinking Encircle and Spiral Update functions. Lastly, the Secant Object Detection function is applied in the third hidden layer to achieve accurate building detection with well-linked edges.

2.6. Development of feature extraction network in SDBN-HCWO model

To enhance feature extraction while minimizing computational costs, this study utilizes a backbone network. The Densely Connected Convolutional Network (DCCN) for SDBN-HCWO method, detailed in Table 2, consists of three key components: 3 × 3 convolutional layers as the initial processor, multiple dense blocks with transition layers for feature extraction and reuse, and a fully connected (FC) layer for classification. Dense blocks serve as the core structural units, ensuring efficient information flow, while transition layers downsample feature maps to improve computational efficiency. Each transition layer includes batch normalization (BN) to stabilize activations, a 1 × 1 convolutional layer for feature refinement, and a 2 × 2 pooling layer for resolution reduction. To optimize detection performance, the dense blocks incorporate batch normalization, the H-Swish activation function, and convolutional layers. By maintaining consistent feature map sizes and interconnecting layers along the channel dimension, the architecture enhances feature propagation, gradient flow, and overall network performance compared to previous deep learning studies.

Unlike ResNet's residual structure, which utilizes skip connections to bypass layers and prevent vanishing gradients, DCCN for SDBN-HCWO method employs a dense connection strategy. In residual networks, only selected layers are connected, whereas DCCN connects every layer to all preceding layers. This design enables each layer to receive inputs from all previous layers and concatenate their outputs along the channel dimension. As a result, it facilitates feature reuse, preserves extracted features, and enhances feature

Table 2
Selection of DCCN parameters.

Layer	Output size	DCCN
Convolution	128 × 128	7 × 7 conv, stride = 2
Pooling	64 × 64	3 × 3 max pool, stride = 2
Dense block-1	64 × 64	6 × (1 × 1 conv, 3 × 3 conv)
Transition Layer-1	64 × 64	1 × 1 conv
	32 × 32	2 × 2 average pool, stride = 2
Dense block-2	32 × 32	12 × (1 × 1conv, 3 × 3conv)
Transition Layer-2	32 × 32	1 × 1 conv
	16 × 16	1 × 1 conv
Dense block-3	16 × 16	24 × (1 × 1conv, 3 × 3conv)
Transition Layer-3	16 × 16	1 × 1 conv
	8 × 8	2 × 2 average pool, stride = 2
Dense block-4	8 × 8	16 × (1 × 1conv, 3 × 3conv)
Classification layer	1 × 1	7 × 7 global average pool
	–	1000D fully-connected, SoftMax

Note: conv is 3 × 3 Convolution.

representation. The dense connections also improve gradient flow during backpropagation, thereby reducing the risk of vanishing gradients. Additionally, this approach promotes parameter efficiency by reusing features instead of adding extra parameters for skip connections. Consequently, DCCN boosts network efficiency while supporting effective feature propagation and learning. The total number of connections in DCCN can be calculated using the following equation:

$$C = 1 + 2 + 3 + \ldots\ldots + (L-1) = \frac{L(L+1)}{2} \tag{15}$$

In this equation, C signifies the total number of connections, while L represents the number of layers in the network.

2.7. Improvement of computational efficiency in SDBN-HCWO model

Depth-wise Separable Convolution (DSC) is a widely recognized technique for improving the computational efficiency of deep neural networks. Originally introduced in MobileNet (Hossain et al., 2021), DSC has since been integrated into various lightweight deep learning architectures. Unlike standard convolution, which applies a single convolutional kernel across all channels of an image (leading to a quadratic increase in parameters and computational cost) DSC significantly reduces these demands while maintaining comparable accuracy (Li et al., 2024). It achieves this by breaking down standard convolution into two steps: depth-wise convolution, where a unique 3 × 3 filter is applied to each channel individually, and point-wise convolution (1 × 1 convolution), which combines feature maps across channels to ensure effective feature extraction. This approach drastically lowers computational expense while preserving essential feature representations. In the SDBN-HCWO detection model, DSC is employed in the neck and head to reduce both computational complexity and the number of parameters. Research indicates that replacing traditional convolutions with DSC can lead to a substantial reduction in computational cost, typically measured by comparing Floating Point Operations per Seconds (FLOPs) before and after its implementation (Langerman et al., 2020). Furthermore, DSC-based optimizations have been shown to enhance real-time performance in object detection models like YOLO, SSD, and EfficientDet, where low-latency inference is critical (Ju et al., 2024). By incorporating DSC into SDBN-HCWO, the model achieves greater efficiency while maintaining detection accuracy, aligning with broader advancements in deep learning optimization. The mathematical expression representing the output of a standard convolution is given as follows:

$$Y_{k,l,n} = \sum_{m}\sum_{i}\sum_{j} K_{i,j,m,n}.F_{k+i-1,l+j-1,m} \tag{16}$$

In the equation: $Y_{k,l,n}$ represents the value at spatial location *(k, l)* in the *n*th channel of the output feature map. $K_{i,j,m,n}$ denotes the convolution kernel parameters, where *i* and *j* represent the spatial positions within the kernel, and mmm corresponds to the input channel index. $F_{k+i-1,l+j-1,m}$ represents the value in the *m*th channel of the input feature map at the corresponding spatial position.

The computational cost expressions for standard convolution and DSC are derived based on the following parameters: Number of input channels (N), Number of output channels (M), Spatial dimensions (width and height) of the input feature map (GF), Spatial dimensions of the convolution kernel (GK). For a standard convolution, the computational cost is determined by the number of floating-point operations (FLOPs) required to compute the output. The expression is:

$$Cos\ t_{std} = M \times N \times GF^2 \times GK^2 \tag{17}$$

DSC breaks the standard convolution into two operations: depth-wise convolution and point-wise convolution (1 × 1 convolution). Each input channel is convolved with its own filter (depth-wise), so the cost for this operation is:

$$Cos\ t_{dw} = N \times GF^2 \times GK^2 \tag{18}$$

After the depth-wise convolution, a 1 × 1 point-wise convolution is applied to combine the outputs of the depth-wise convolutions. The cost for the point-wise convolution is:

$$Cos\ t_{pw} = N \times M \times GF^2 \tag{19}$$

Thus, the total computational cost for DSC is:

$$Cos\ t_{DSC} = Cos\ t_{dw} + Cos\ t_{pw} = \left(N \times GF^2 \times GK^2\right) + \left(N \times M \times GF^2\right) \tag{20}$$

$$\frac{Cos\ t_{DSC}}{Cos\ t_{std}} = \frac{1}{M} + \frac{1}{GK^2} \tag{21}$$

As shown in the equations, DSC reduces the computational cost compared to standard convolution, especially when the number of output channels M is large.

The mathematical expression for the *m*th channel of the output feature map in a DSC can be written as:

$$\widehat{Y}_{k,l,n} = \sum_{i}\sum_{j} \widehat{K}_{i,j,m}.F_{k+i-1,l+j-1,m} \tag{22}$$

In the equation: $\widehat{Y}_{k,l,n}$ represents the value at spatial position *(k, l)* in the *m*th channel of the output feature map. $\widehat{K}_{i,j,m}$ denotes the depth-wise convolution kernel parameters, where *(i, j)* represent the spatial positions within the kernel. The depth-wise separable

convolution mechanism is shown in Figs. 5 and 6.

2.8. Improve detection performance in SDBN-HCWO model

In deep learning-based object detection, selecting the right activation function is crucial for model accuracy and efficiency. Research shows that activation functions significantly impact feature extraction, gradient flow, and overall model convergence. YOLOv4, a popular object detection model, uses the Conv2d Leaky ReLU activation function in its backbone network to tackle the issue of neuron death (Shi et al., 2021). By permitting a small gradient for negative input values, Leaky ReLU helps ensure continuous learning during backpropagation. However, studies have revealed that its performance can vary across different data distributions, leading to unreliable feature representations and limiting the model's ability to generalize effectively. To address these limitations in detecting building, the proposed SDBN-HCWO model incorporates H-Swish as its primary activation function. H-Swish, a refined version of Swish, is known for its smooth and continuous characteristics, which enhance gradient propagation and computational efficiency. Unlike Leaky ReLU, which follows a fixed piecewise linear approach, H-Swish features a self-gating mechanism that dynamically adjusts activation values, thereby improving network stability and representational power. Previous research has demonstrated its effectiveness, especially in lightweight models, making it a suitable choice for real-time object detection applications. The implementation of H-Swish in the proposed model aims to enhance detection accuracy and reliability across various conditions, as illustrated in the comparative function graphs in Fig. 8. The mathematical expression for the Leaky ReLU activation function is:

$$Leaky\ ReLU(x) = \begin{cases} x & if\ x \geq 0 \\ \alpha x & if\ x < 0 \end{cases} \tag{23}$$

where, x represents the input value to the activation function. α is a small positive constant (typically 0.01), allowing a slight gradient for negative input values to prevent neuron death. The mathematical expression for the H-Swish (Hard-Swish) activation function is:

$$H - Swish(x) = x\frac{\mathrm{Re}LU6(x+3)}{6} \quad Here\ \mathrm{Re}LU6(x) = \min[\max(0,x),6] \tag{24}$$

The flow-diagram of DSC in SDBN-HCWO model for computational efficiency is shown in Fig. 7.

The H-Swish activation function offers numerous benefits, making it an appealing choice for deep learning models, particularly in object detection tasks. Like ReLU, it does not have an upper limit, which helps avoid gradient saturation—a common problem where gradients become too small to propagate effectively, thereby slowing down the model's learning process. This characteristic enables the model to continue learning efficiently, preventing training slowdowns. Additionally, H-Swish has a lower bound that causes its values to approach zero on the negative side. This property helps prevent overfitting by promoting more balanced weight updates, particularly when training with extreme values. Another significant advantage of H-Swish is its non-monotonicity, meaning it can retain small negative values instead of zeroing them out. This feature contributes to a more stable flow of gradients throughout the network, which is vital for consistent learning. In contrast to standard activation functions like ReLU, which discard negative values and risk leaving neurons inactive, H-Swish maintains these small negative values, ensuring proper weight adjustments. Furthermore, the function is continuous and differentiable, allowing for smoother optimization and more efficient weight updates, which enhances the convergence process. Ultimately, H-Swish provides excellent generalization and optimization capabilities, significantly improving model performance and recognition accuracy in tasks such as object detection. These qualities highlight its effectiveness in enhancing both the accuracy and training stability of neural networks.

In conclusion, the H-Swish activation function enhances detection precision, particularly for identifying damaged illegal construction objects. Its non-monotonicity prevents neuron saturation, promoting stable learning and improving model accuracy. Additionally, the lower bound helps reduce overfitting by limiting extreme values during training. H-Swish also offers improved computational efficiency over the Swish function, accelerating training and minimizing resource consumption. This makes it suitable for real-time applications that require rapid processing and low latency. Overall, H-Swish is an effective activation function for enhancing both the accuracy and efficiency of detection models.

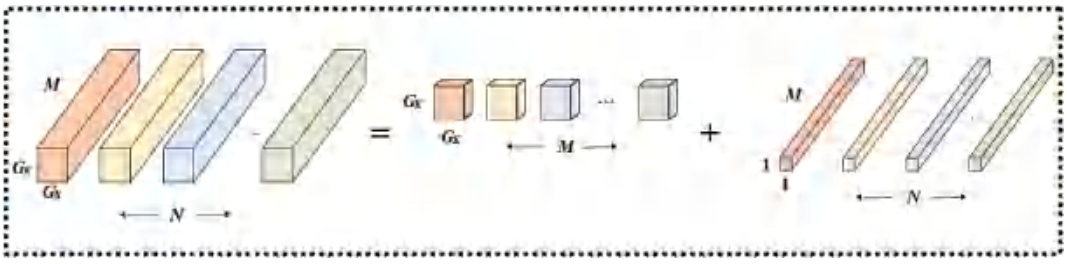

Fig. 5. Framework of depth-wise separable convolution mechanism.

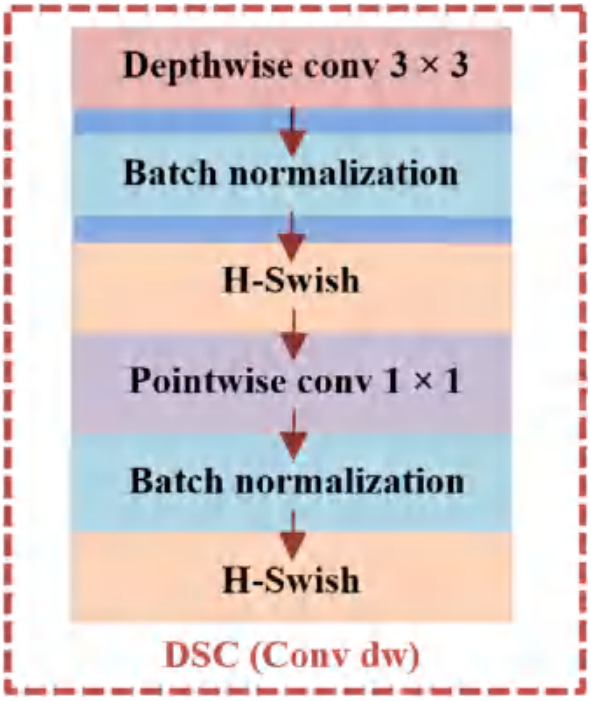

Fig. 6. Mechanism of DSC for proposed SDBN-HCWO.

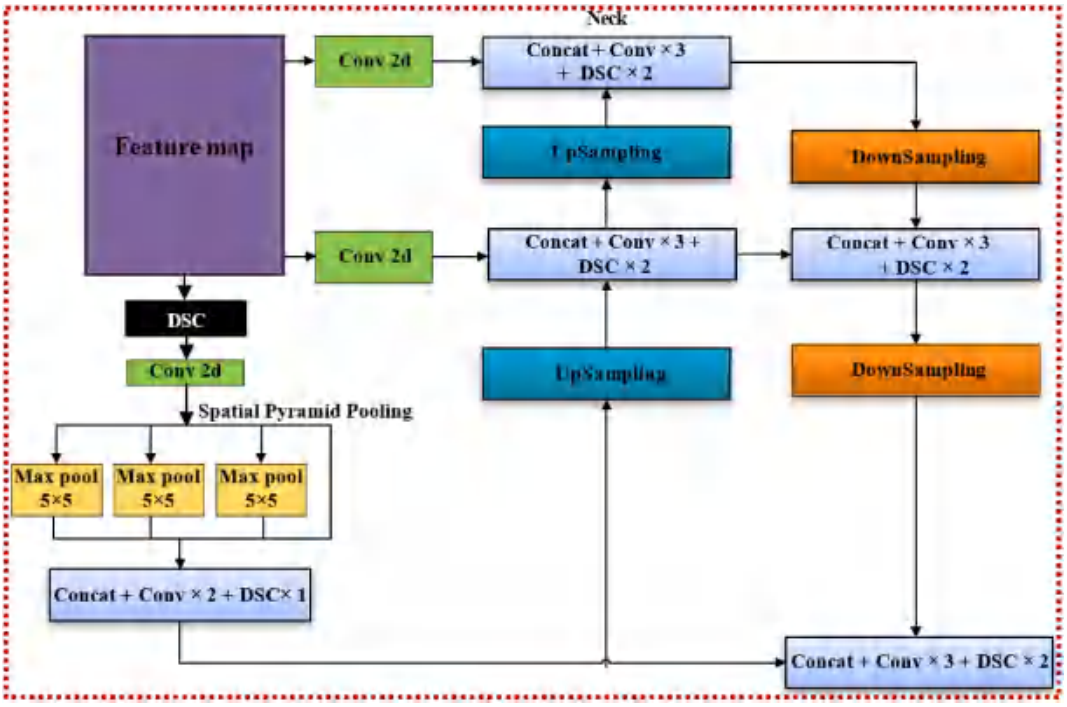

Fig. 7. Functional flow-diagram of DSC in SDBN-HCWO model for computational efficiency.

2.9. Developed high accuracy object detection model with SDBN-HCWO

Fig. 9, presents the architecture of the proposed SDBN-HCWO detection model, which outperforms YOLOv4 in accuracy and parameter efficiency. Replacing CSPDarkNet53 (Senussi and Kang, 2024) with DCCN improves feature extraction while maintaining

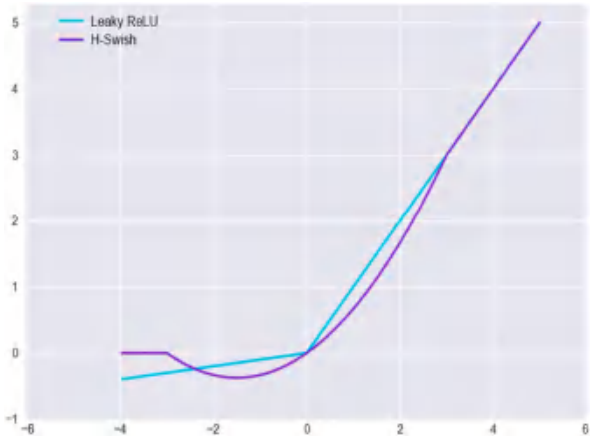

Fig. 8. Graphical change of the H-Swish and Leaky ReLU activation functions.

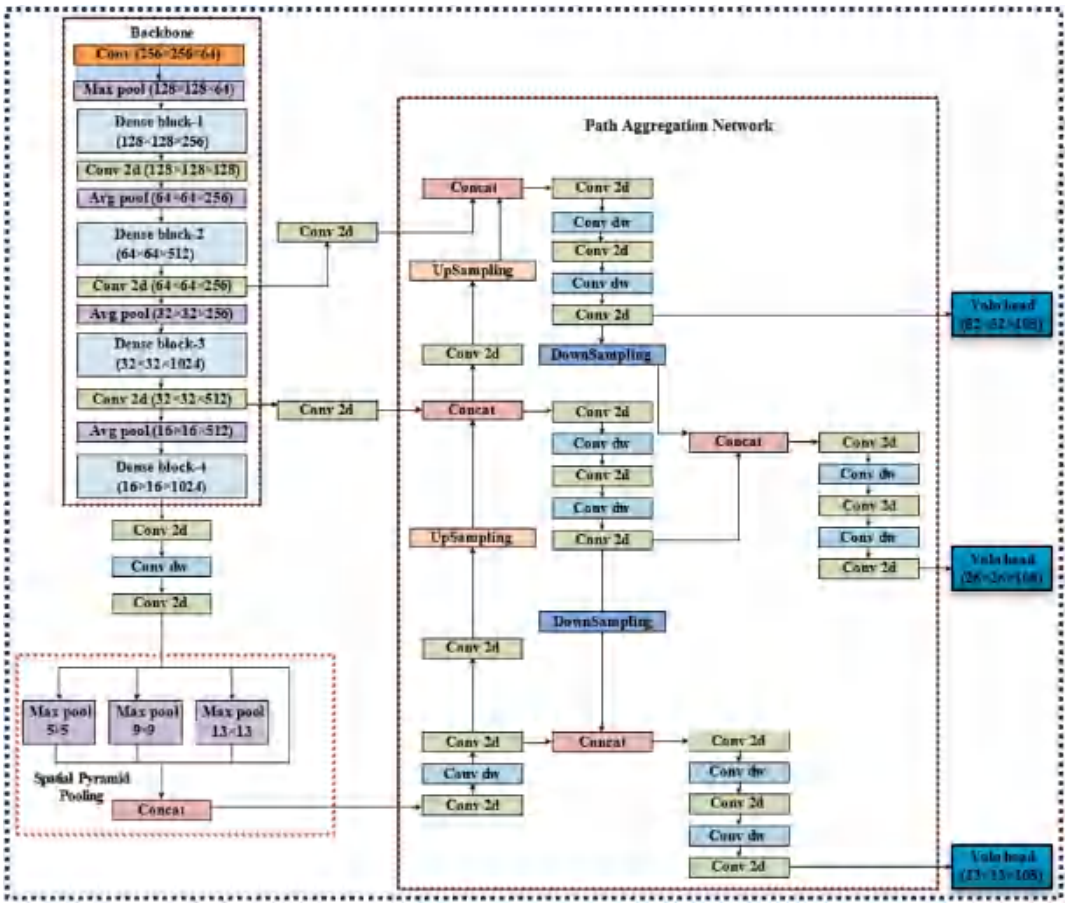

Fig. 9. Architectural network of proposed building detection model with SDBN-HCWO model.

key network components like the 3 × 3 convolutional layer, pooling layers, and dense blocks. DCCN enhances feature transfer, mitigates gradient vanishing, and optimizes feature utilization, boosting recognition accuracy. While feature layer channels are adjusted, their size remains unchanged, and dense connections reduce parameter count, improving efficiency. Further enhancements include replacing the standard 3 × 3 convolution in the neck and head with depth wise separable convolution (DSC) to lower computational costs. H-Swish, a more effective activation function than Leaky ReLU, improves nonlinear feature extraction and generalization. Experiments confirm that these modifications enhance detection accuracy while reducing computational costs, making

SDBN-HCWO superior to other state-of-the-art (SOTA) models for illegal construction detection.

The loss function of SDBN-HCWO consists of multiple key components, including the object position loss function (L_l), classification loss function (L_c), confidence loss function (L_{cf}), *IoU* loss (L_{iou}), scale-invariant loss (L_s), and edge loss (L_{edge}). The overall loss function can be expressed as:

$$L = \lambda_l L_l + \lambda_c L_c + \lambda_{cf} L_{cf} + \lambda_{iou} L_{iou} + \lambda_s L_s + \lambda_{edge} L_{edge} \tag{25}$$

Where, $\lambda_l, \lambda_c, \lambda_{cf}, \lambda_{iou}, \lambda_s, \lambda_{edge}$ are weighting factors to controls the relative importance of its corresponding loss component.

$$L_l = 1 - \frac{B \cap \widehat{B}}{B \cup \widehat{B}} \tag{26}$$

Where, $B \cap \widehat{B}$ is the overlapping area between the predicted and ground truth bounding boxes. $B \cup \widehat{B}$ is the total area covered by both boxes.

$$L_c = \sum_{i=1}^{N} y_i \log(\widehat{y}_i) + (1 - y_i)\log(1 - \widehat{y}_i) \tag{27}$$

Where, N is the number of samples. y_i is the ground truth label (1 if the object belongs to the positive class, 0 if it belongs to the negative class). $\widehat{y}_i$ is the predicted probability of the object belonging to the positive class.

$$L_{cf} = -\sum_{i=1}^{N} c_i \log(\widehat{c}_i) + (1 - c_i)\log(1 - \widehat{c}_i) \tag{28}$$

Where, c_i is the ground truth confidence (1 for object present, 0 for no object). $\widehat{c}_i$ is the predicted confidence score for bounding box *N*, representing the probability that an object is detected in the box.

$$L_l(IoU) = \begin{cases} 0.5(IoU - 1)^2 & if\ |(IoU - 1)| < 1 \\ |(IoU - 1)| - 0.5 & Otherwise \end{cases} \tag{29}$$

If *IoU* is close to 1 (i.e., the predicted box is nearly identical to the ground truth), the loss is small. If *IoU* is far from 1 (i.e., the predicted box does not overlap well with the ground truth), the loss is larger.

$$L_{scale} = \frac{1}{2}\left(\frac{|w_{pred} - w_{gt}|}{w_{pred} + w_{gt}} + \frac{|h_{pred} - h_{gt}|}{h_{pred} + h_{gt}}\right) \tag{30}$$

Where, $|w_{pred} - w_{gt}|$ is the absolute difference in width between the predicted and ground truth bounding boxes. $|h_{pred} - h_{gt}|$ is the absolute difference in height between the predicted and ground truth bounding boxes. The sum $w_{pred} + w_{gt}$ and $h_{pred} + h_{gt}$ in the denominator normalizes the loss, making it scale-invariant.

$$L_{edge} = \frac{1}{N}\sum_{i=1}^{N}\left[\left|E_{pred}(i) - E_{gt}(i)\right|^2\right] \tag{31}$$

Where, $E_{pred}(i)$ represents the predicted edge at pixel *i*. $E_{gt}(i)$ represents the ground truth edge at pixel *i*. *N* is the total number of pixels in the boundary region.

3. Experiments and results

3.1. Experimental setup

The proposed SDBN-HCWO method, along with the existing Mask R-CNN and Dilated ResUnet deep learning models, is evaluated using Python to assess its effectiveness in recognizing buildings from satellite images. These building detection methods utilize the Massachusetts Buildings Dataset, which is available on Kaggle (Massachusetts Buildings Dataset). The Massachusetts Buildings Dataset comprises 151 aerial images capturing the Boston area, each with a resolution of 1500 × 1500 pixels, representing an area of 2.25 square kilometres per image. Collectively, the dataset spans approximately 340 square kilometres. It is divided into 1137 images for training, 216 for testing, and 104 for validation. The building footprint maps were created by rasterizing data sourced from the OpenStreetMap project, focusing on areas with an omission noise level of around 5 % or lower. The dataset benefits from high-quality building footprint data, largely due to the City of Boston's contribution of comprehensive building outlines to OpenStreetMap. It primarily covers urban and suburban regions and includes buildings of various sizes, such as houses and garages. The imagery, provided by the state of Massachusetts, is standardized to a resolution of 1 pixel per square meter. To enhance evaluation accuracy, the target maps for the test and validation sets underwent manual corrections.

To evaluate the effectiveness of the SDBN-HCWO method, this research conducted a comparative analysis against the Mask R-CNN and Dilated-ResUnet models. This comparison utilized various performance metrics, including PSNR, false positive rate, Classification

Accuracy (CA), and Confusion Matrix (CM). These metrics provide a comprehensive evaluation of each method's performance in building detection tasks. The performance of building detection using the proposed SDBN-HCWO method is evaluated and compared with traditional building detection approaches. Specifically, it is compared against the Mask R-CNN developed by Han et al. and the Dilated-ResUnet deep learning model introduced by Dixit et al. (Dixit et al., 2021; Han et al., 2022). The comparison is based on key performance metrics, including PSNR, false positive rate, classification accuracy, and classification time. To ensure a fair and consistent evaluation, we used 130 diverse image samples as input across all three methods. Additionally, an average of 10 simulation runs was conducted to account for variability and enhance the reliability of the results. This comparison provides a comprehensive assessment of the effectiveness and efficiency of the SDBN-HCWO method in building detection tasks.

In addition, the SDBN-HCWO model is used in various scenarios, such as damage detection in natural disasters (Deep Learning for Satellite Image Processing), including identifying buildings in initial images and assessing the extent of damage to each building after a disaster. The dataset comprises 9168 samples, including 3240 images that depict damaged structures. Due to research constraints, only a subset of the dataset is used for training: 810 (25 % total datasets) images are allocated for training, 324 (10 % total datasets) for validation, and 162 (5 % total datasets) as testing samples. Although this limited sample size is not sufficient for fully training a segmentation model, it offers a controlled setting to evaluate the SDBN-HCWO methodologies. In the experimental setup, the dataset is categorized into four distinct classes, ensuring a multi-class classification framework suited for object detection or recognition tasks. The model is trained over 100 epochs, allowing sufficient iterations for learning complex patterns within the data. A batch size of 4 is employed, which facilitates frequent weight updates and enhances learning stability, particularly when computational resources are limited or when dealing with high-resolution input data. Additionally, a learning rate of 0.001 is adopted, striking a balance between convergence speed and training stability. This learning rate enables the model to make gradual yet effective adjustments to its parameters, reducing the risk of overshooting the optimal solution during training.

Data augmentation is a crucial technique in deep learning, particularly for image-based tasks, as it enhances model generalization by artificially expanding the training dataset. Two commonly used augmentation methods are flipping and rotation. Flipping involves horizontally or vertically inverting an image, which allows the model to learn invariant features despite the orientation. Rotation, on the other hand, involves turning the image by a certain angle, such as 90°, 180°, or random degrees, which further exposes the model to various spatial representations of the same object. The dataset is processed under consistent conditions to ensure reproducibility, utilizing a fixed random seed and deterministic GPU execution mode. For a thorough analysis, using the full dataset is recommended to validate the scalability of the findings. To aid in defect identification, the dataset is categorized into four groups, as outlined in Table 3.

3.2. Visual display of the experiments

Fig. 10(a)–(d) present the input edge-linked images, which highlight the structural outlines and contours within the scenes, serving as a crucial pre-processing step for enhancing feature extraction. These edge-linked representations help in emphasizing the geometric boundaries of objects, thereby facilitating more accurate localization in the subsequent detection phase. Correspondingly, Fig. 10(e)–(h) illustrate the output results after applying the SDBN-HCWO. The model successfully identifies and localizes the target objects within each scene, demonstrating its effectiveness in interpreting edge-based input data and translating it into accurate detection outputs. This comparison between input and output visually underscores the model's capability in transforming low-level edge information into meaningful object detection results.

3.3. Performance of PSNR

The Peak Signal-to-Noise Ratio (PSNR) is calculated using the mean square error (MSE), which measures the difference between the original and processed images. PSNR quantifies the quality difference between these two images, with higher values indicating better image quality and less distortion. Mathematically, PSNR is defined as a logarithmic function of the MSE, creating a scale for evaluating the effectiveness of image processing methods. This metric is widely used in tasks such as image compression, restoration, and detection to assess how closely the processed image resembles the original. The calculation of PSNR is given by the following formula (Kononchuk et al., 2022; Basha and Logu, 2024):

$$PSNR = 10 \times \log_{10}\left[\frac{M^2}{S_{ME}}\right] \tag{32}$$

$$S_{ME} = \left[Size_{preprocessed} - Size_{original}\right]^2 \tag{33}$$

From equations (15) and (16), the term $Size_{preprocessed}$ refers to the size of the processed image, specifically the best positive vector,

Table 3
Category and definition of building damage assessment task.

No.	Category	Definition
1	Destroyed Buildings	The damage is irreversible, with major collapses, severe cracks, or total destruction of walls, roofs, and foundations.
2	Light Damaged Buildings	Damage may include small cracks, broken windows, minor roof or wall damage, and cosmetic issues.
3	Medium Damaged Buildings	Damage includes large cracks, partial wall collapses, deformed structures, and weakened load-bearing elements.
4	Non-Damaged Buildings	These structures have no visible damage and remain completely functional.

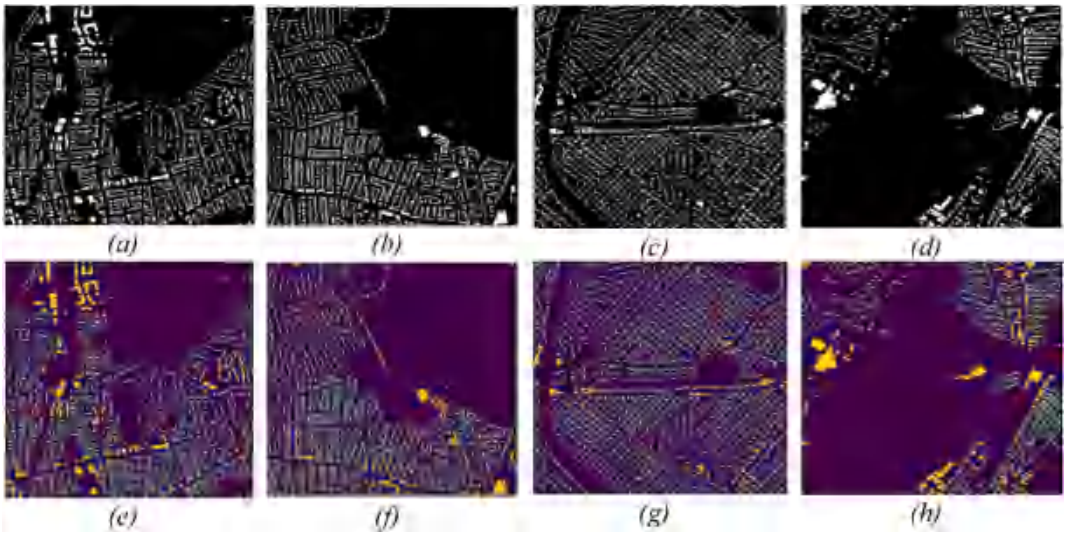

Fig. 10. (a), (b), (c), (d) – edge linked images (e), (f), (g), (h) – Output result after applying proposed SDBN-HCWO.

while $Size_{original}$ denotes the size of the original image. PSNR, or Peak Signal-to-Noise Ratio, is measured in decibels (dB) and provides a quantitative assessment of image quality by comparing original and processed images. A higher PSNR value indicates better image quality with less distortion. The results of the comparative analysis of PSNR for the proposed method and existing methods are presented in Fig. 11. This figure offers a clear comparison of the performance of each approach in terms of image quality.

Fig. 11 illustrates the PSNR performance for various image sizes derived from the Massachusetts Buildings dataset. The PSNR values achieved using the proposed SDBN-HCWO method are compared with those from existing methods, such as Mask R-CNN and Dilated-ResUnet deep learning models, to assess the effectiveness of building detection. The comparison clearly shows that the SDBN-HCWO method outperforms the existing methods in terms of PSNR, indicating superior image quality and detection accuracy in building recognition tasks.

Fig. 11 presents the performance analysis of PSNR for the proposed SDBN-HCWO method in comparison to existing techniques. The x-axis represents the varying image sizes measured in megabytes (MB), while the y-axis displays the corresponding PSNR values for each method. The results show that PSNR remains relatively stable across all three methods for images ranging from 4.34 MB to 4.93 MB, followed by a gradual decline and then an increase. This fluctuation in PSNR is attributed to the presence of occlusion in the images. Specifically, simulations with a 4.34 MB image size yield a PSNR of 39.75 dB for the proposed SDBN-HCWO method, 37.25 dB for the Mask R-CNN, and 33.73 dB for the Dilated-ResUnet deep learning method. Notably, the SDBN-HCWO method consistently achieves higher PSNR values, indicating superior performance in building detection compared to the existing methods.

The PSNR achieved by the proposed SDBN-HCWO method can be attributed to the integration of the Hyperbolic Cosine function. This function is essential for reducing the risk of position updates in the best edge selection process from becoming trapped in local

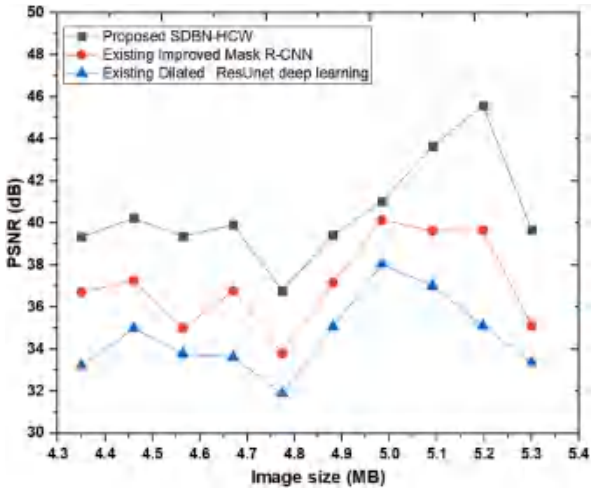

Fig. 11. Results PSNR using SDBN-HCWO, Mask R-CNN and Dilated-ResUnet deep Learning.

optima, which is a common issue in optimization. To tackle this challenge, a decreasing vector is used to effectively manage local optimal solutions. The Hyperbolic Cosine function facilitates a balanced process of exploration and exploitation, contributing to the enhancement of PSNR in the SDBN-HCWO method. Simulation results indicate that the proposed method improves PSNR by 9 % compared to the Mask R-CNN and by 17 % compared to the Dilated-ResUnet deep learning approach, highlighting its superior performance in edge detection.

3.4. Performance of FPR

The False Positive Rate (FPR) is calculated by examining the number of negative instances. Specifically, the objects that are not classified as building pixels but are incorrectly identified as positive (i.e., building pixels). To compute the FPR, you divide the number of false positives (FP) by the total number of actual negative instances (TN). This measurement indicates how often non-building pixels are mistakenly classified as building pixels. The formula for estimating the FPR is as follows (Dhanaraj et al., 2021; Balyan et al., 2022):

$$FPR = \frac{FP}{FP + TN} \tag{34}$$

In Equation (17), 'FP' refers to false positives, while 'TN' stands for true negatives. FPR quantifies the proportion of incorrectly identified positive instances compared to the total number of actual negative cases. Fig. 12 provides a comprehensive summary of the varying FPR results, showcasing the performance of the different methods under consideration. This figure highlights the differences in accuracy and error rates across the models, offering valuable insights for comparison.

Specifically, Fig. 12 displays the FPR results derived from a diverse set of sample images taken from the Massachusetts Buildings Dataset. In the simulations, a total of 130 distinct images, ranging in size from 13 to 130, were utilized. All three methods effectively reduced the FPR in the building detection task. However, the proposed SDBN-HCWO method demonstrated superior performance by achieving a lower FPR compared to the existing Mask R-CNN and Dilated-ResUnet deep learning methods. This indicates that the SDBN-HCWO approach enhances accuracy by minimizing false positives during building detection.

Fig. 12 presents the simulation output of the FPR for 130 distinct input images. As shown in the figure, the FPR increases with the number of input images, which can be attributed to the varying image sizes and corresponding noise levels. This variation in noise results in a proportional increase in falsifications during the edge linking process used for building detection. Despite this challenge, the simulations conducted across various image sizes demonstrate that the proposed SDBN-HCWO method outperforms other methods by achieving the lowest FPR. This improvement is largely due to the application of the Spiral Updating Equation, which incorporates an adjustment coefficient during the exploitation phase. The equation resembles a helix-shaped series of edges, effectively linking the best edges while eliminating isolated ones. As a result, the FPR of the SDBN-HCWO method is reduced by 44 % compared to the Mask R-CNN and by 64 % compared to the Dilated-ResUnet deep learning methods.

3.5. Performance of classification accuracy

Classification accuracy (CA) is an important performance metric used in building detection tasks. It quantifies how effectively a detection method identifies buildings by comparing the number of correctly classified building pixels to the total number of pixels, including both buildings and non-buildings. This measure is essential for evaluating the reliability and precision of detection models, as it offers insights into the overall accuracy of the classification process. By assessing CA, researchers can determine how well proposed methods distinguish between relevant and irrelevant features, thereby ensuring the quality and efficiency of the building

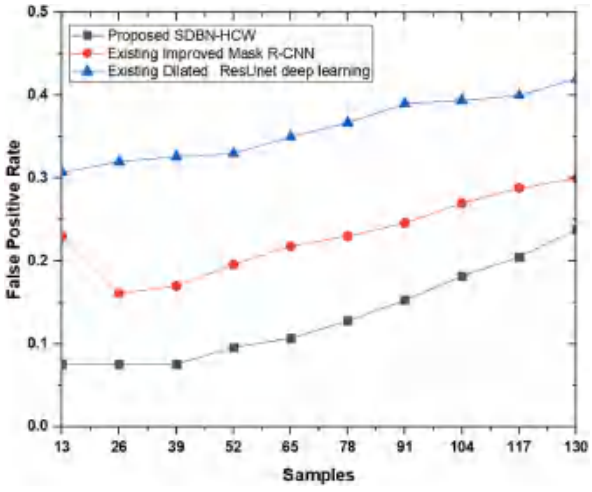

Fig. 12. Results FPR using SDBN-HCWO, Mask R-CNN and Dilated-ResUnet deep learning.

detection system (Javed et al., 2021; Uddin et al., 2022).

$$CA = \sum_{i=1}^{n} \frac{Samples_{cc}}{Samples_i} \times 100 \tag{35}$$

In equation (18), "CA" stands for Classification Accuracy, which measures the proportion of correctly identified building objects within a set of images. The variable "$Samples_i$" refers to the total number of sample images used for evaluation, while "$Samples_{CC}$" indicates the number of sample images that were correctly classified, representing the building objects accurately detected as buildings. Classification accuracy is expressed as a percentage (%), providing a clear insight into the model's performance.

Fig. 13 presents a detailed breakdown of classification accuracy for 130 distinct sample images, enabling a comparative analysis of the proposed method against other detection techniques. The experimental results shown in Fig. 13 illustrate the CA across various sample sizes, ranging from 13 to 130 images. During the simulation, the CA performance of the proposed SDBN-HCWO method was compared to that of existing methods. The results indicate a significant improvement in classification accuracy when utilizing the SDBN-HCWO approach. Specifically, this proposed method consistently outperformed the existing techniques, demonstrating a marked increase in the accuracy of building detection. This improvement underscores the effectiveness of the SDBN-HCWO method in enhancing classification performance compared to traditional approaches.

Fig. 13 presents the classification accuracy results for the proposed SDBN-HCWO method, along with the Mask R-CNN and Dilated-ResUnet deep learning approaches. The horizontal axis indicates the number of sample images used in the simulation, while the vertical axis displays the corresponding classification accuracy. The figure shows a slight decline in classification accuracy as the number of input images increases. This decrease is due to the difficulties in accurately identifying and linking edges as the sample size grows, which can lead to a reduction in the accuracy of building detection. Despite this trend, the SDBN-HCWO method achieved a classification accuracy of 92.3 % when tested with 26 sample images, surpassing the Mask R-CNN (76.92 %) and Dilated-ResUnet (69.23 %) techniques. These results highlight the superior performance of the SDBN-HCWO method in maintaining high accuracy across varying sample sizes.

Additionally, the Secant Object Detection model integrated into the SDBN-HCWO method significantly contributes to achieving higher classification accuracy. This model effectively explores the entire search space by expanding from the current best solution to identify optimal choices. By applying the secant equation, the method utilizes the proximity of the target to the current estimate, ensuring precise building detection. This approach guarantees that the curvature constraints encompass the entire image within the visible layer. As a result, the SDBN-HCWO method demonstrates a 12 % improvement in classification accuracy compared to the Mask R-CNN and a significant 35 % increase over the Dilated-ResUnet method. These findings underscore the superior performance of the SDBN-HCWO method in accurately classifying sample images.

3.6. Performance of CT

The time taken for image classification in building detection is analysed. This time, referred to as Classification Time (CT), measures the duration needed to execute three distinct processes across three hidden layers involved in building detection. The classification time is expressed in milliseconds (*ms*), providing a precise quantification of the computational effort required to categorize images. This measurement is essential for evaluating the efficiency of the proposed method, ensuring that the classification process is both accurate and time-efficient (Sanhudo et al., 2021; Pin Tan et al., 2021).

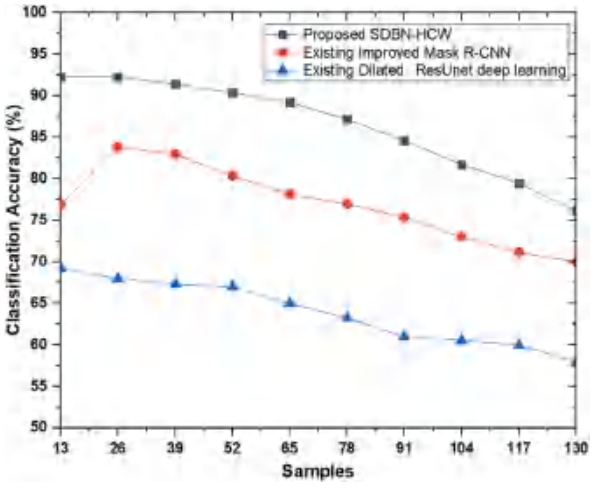

Fig. 13. Results classification accuracy using SDBN-HCWO, Mask R-CNN and Dilated-ResUnet deep learning.

$$CT = \sum_{i=1}^{n} Samples_i \times Time[\text{Res}] \tag{36}$$

In equation (19), '*CT*' represents the classification time, while '*Time* [*Res*]' refers to the time required to identify the building within the image.

Fig. 14 illustrates the simulation results for CT across various sample sizes. During the simulation, the classification time performance of the proposed SDBN-HCWO method is compared with that of other existing methods. The data indicates a significant reduction in classification time when using the SDBN-HCWO method, highlighting its efficiency in building detection. Compared to existing methods, the SDBN-HCWO approach demonstrates a more time-efficient process, thereby minimizing the overall computational time needed for accurate building detection. This reduction in classification time underscores the practical advantages of the proposed method in real-world applications.

In Fig. 14, the classification time results for the SDBN-HCWO method, Mask R-CNN, and Dilated-ResUnet deep learning approach are presented. The graph displays classification time on the y-axis and the number of sample images on the x-axis. As the number of images increases, the building detection performance improves, leading to a corresponding rise in classification time. However, simulations conducted with 13 sample images reveal that the SDBN-HCWO method achieves a classification time of 4.94 ms, which is faster than the 6.89 ms for Mask R-CNN and 7.63 ms for the Dilated-ResUnet deep learning method. This analysis demonstrates that the proposed SDBN-HCWO method is more time-efficient in building detection than existing methods, offering faster classification times while maintaining performance.

The reduction in classification time can be attributed to the implementation of the SDBN-HCWO algorithm. In the first hidden layer, the Hyperbolic Cosine Prey Encircling function is applied to detect edges. In the second hidden layer, the edges are linked by fine-tuning the initialized coefficient vectors using the shrinking encircle and spiral update functions. Finally, building detection is performed through the Secant Object Detection function, which utilizes the linked edges. This approach effectively minimizes classification time in the proposed SDBN-HCWO method, achieving a 31 % reduction compared to the Mask R-CNN and a 39 % reduction compared to the Dilated-ResUnet deep learning methods.

3.7. Validation of SDBN-HCWO with different scenario

To assess the robust effectiveness of the proposed SDBN-HCWO building detection model, widely used metrics in object detection were selected for evaluation. These include Average Precision (AP), precision, recall, and F1-score. The corresponding calculation formulas for these metrics are presented as follows:

$$\text{Precision Rate} = \frac{TP}{TP + FP} \tag{37}$$

$$\text{Recall Rate} = \frac{TP}{TP + FN} \tag{38}$$

$$\text{F1} - \text{Score} = \frac{2TP}{2TP + FP + FN} \tag{39}$$

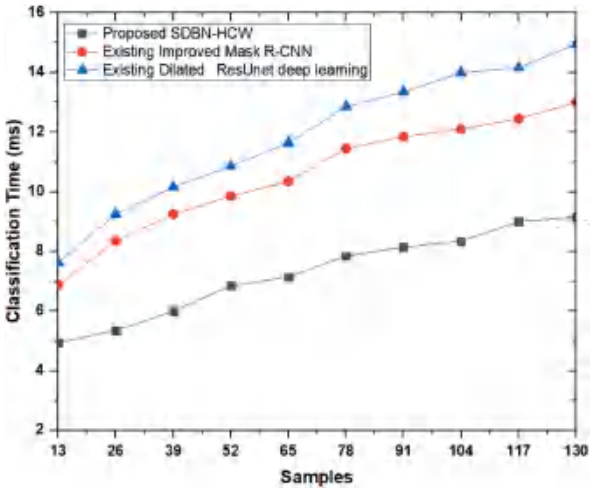

Fig. 14. Results classification time using SDBN-HCWO, Mask R-CNN and Dilated ResUnet deep learning.

$$AP = \int_0^1 P(R)dR \tag{40}$$

True Positives (TP) denote the correctly identified objects. False Positives (FP) refer to objects that are incorrectly classified. False Negatives (FN) represent objects that remain undetected.

As shown in Table 4, the SDBN-HCWO model demonstrates strong performance in detecting building damage across various categories. The model excels at identifying destroyed buildings, achieving the highest Recall rate of 92.24 %, which ensures that severely damaged structures are accurately detected. However, a slightly lower Precision of 87.33 % indicates some instances of misclassification. When it comes to detecting lightly damaged buildings, the model maintains balanced performance, with an AP of 88.13 %, Precision of 88.53 %, and Recall of 88.51 %, resulting in a reliable F1-score of 88.71 %. The detection of medium damage presents notable challenges, as reflected by the category's lowest Average Precision (AP) of 78.51 % among all damage classes. AP is a comprehensive performance metric that combines both Precision and Recall across various confidence thresholds, and a lower AP indicates difficulty in achieving a consistent balance between correctly identifying true positives and avoiding false positives. Interestingly, despite the low AP, the medium damage category achieves the highest Precision value of 96.31 %, meaning that when the model predicts a building to be medium-damaged, it is correct in the vast majority of cases. This high Precision implies that the model is conservative in its predictions, only labelling buildings as medium-damaged when it is highly confident. However, this conservative approach results in a trade-off, as evidenced by a relatively lower Recall of 79.00 %. This means that the model fails to detect approximately 21 % of actual medium-damaged buildings, indicating that some true cases are missed. The combination of high Precision and lower Recall suggests that while the model is accurate when it does predict medium damage, it may overlook less obvious or borderline cases, highlighting the need for further refinement in recognizing the more subtle features associated with medium structural damage. Conversely, non-damaged buildings are identified with high accuracy, evidenced by an AP of 92.67 % and consistent Precision, Recall, and F1-score of 90.67 %. Overall, the model achieves an average AP of 87.43 %, with strong Precision at 90.71 % and an F1-score of 88.96 %, indicating effective detection capabilities. While the model performs well in recognizing both severely damaged and non-damaged buildings, there is room for improvement in the detection of medium-damaged structures to enhance overall classification accuracy.

Table 5 presents an analysis of the SDBN-HCWO model's performance in detecting building damage across various disaster-affected locations. The model demonstrates high accuracy, especially in scenarios involving hurricanes and floods. Among the evaluated locations, the Nepal flooding incident achieved the highest Average Precision (AP) of 90.24 %, with strong scores in precision (89.17 %) and recall (88.16 %), resulting in a solid F1-score of 87.69 %. Similarly, Hurricane Florence exhibited commendable performance, maintaining a well-balanced precision-recall trade-off with an AP of 90.22 %. Overall, locations affected by hurricanes (including Harvey, Florence, and Matthew) along with the Palu tsunami and Nepal flooding consistently achieved AP scores above 88 %, indicating the model's robustness in detecting damage. Notably, Hurricane Harvey recorded the highest precision at 90.53 %, suggesting a minimal rate of false positives. However, some locations, like Joplin (tornado) and Hurricane Matthew, exhibited lower recall values (81.84 % and 81.91 % respectively), indicating a higher likelihood of missed detections. The Guatemala volcano event recorded the lowest F1-score at 83.64 %, reflecting an imbalance between precision and recall. These findings indicate that while SDBN-HCWO is highly effective in damage detection, variations in recall highlight the challenges different disaster environments pose, with complex structural damage patterns affecting detection accuracy. Nevertheless, the model maintains strong overall performance, particularly in regions affected by hurricanes and floods.

The comparative analysis presented in Table 6 shows that the SDBN-HCWO model outperforms its competitors across several key evaluation metrics. With an AP of 87 %, the SDBN-HCWO model surpasses all others, significantly outperforming YOLOv3 and SSD, both of which achieved an AP of 80 %. In terms of precision, SDBN-HCWO achieves the highest score at 91 %, again exceeding YOLOv3 (84 %) and SSD (83 %). For recall, SDBN-HCWO leads with a score of 88 %, followed by YOLOv4 (81 %) and YOLOv3 (80 %). Additionally, it boasts the best F1-score at 89 %, greatly surpassing both YOLOv3 and YOLOv4, which each scored 81 %. Although the SDBN-HCWO model is not the smallest in terms of parameter size, weighing in at 17.2 MB, it strikes an optimal balance between efficiency and accuracy. It remains significantly smaller than Faster R-CNN (137.30 MB), YOLOv3 (61.69 MB), and YOLOv4 (64.10 MB), while still delivering superior performance. Overall, these findings establish the SDBN-HCWO model as a highly efficient and accurate detector, outperforming existing models in both detection capability and computational efficiency.

The performance of the SDBN-HCWO model was evaluated under different configurations, as shown in Table 7. This evaluation involved integrating DenseNet121, Depth-wise Separable Convolution (DSC), and the H-Swish activation function. The assessment metrics illustrate the effects of these enhancements on detection accuracy. The baseline SDBN-HCWO model, which did not include

Table 4
Performance analysis of SDBN-HCWO for detecting damage in buildings.

Category	AP (%)	Precision (%)	Recall (%)	F1-score (%)
Destroyed buildings	90.40	87.33	92.24	89.71
Light damaged buildings	88.13	88.53	88.51	88.71
Medium damaged buildings	78.51	96.31	79.00	86.76
Non damaged buildings	92.67	90.67	90.67	90.67
Total Building Damage	87.19	91.44	87.30	89.05
Average	87.4275	90.71	87.605	88.9625

Table 5
Analysis of the performance of SDBN-HCWO in detecting damage in buildings across various locations.

Place Name	AP (%)	Precision (%)	Recall (%)	F1-score (%)
Guatemala-volcano	86.27	86.31	83.68	83.64
Woolsey-fire	89.04	88.26	85.04	86.79
Joplin-tornado	87.53	87.35	81.84	83.25
Hurricane-Matthew	88.81	88.23	81.91	83.86
Hurricane-Harvey	89.48	90.53	87.62	88.94
Hurricane-Florence	90.22	88.48	87.92	87.29
Palu-Tsunami	88.48	88.59	86.15	87.95
Nepal-Flooding	90.24	89.17	88.16	87.69

Table 6
Comparing the performance of the proposed SDBN-HCWO model with existing model.

Detector	AP (%)	Precision (%)	Recall (%)	F1-score (%)	Size of parameters (MB)
SSD (Li et al., 2019)	80	83	78	79	27.62
RetinaNet (Walter et al., 2023)	79	74	79	76	36.95
Faster R-CNN (Ding et al., 2022)	73	53	84	64	137.30
YOLOv5-L (Yang et al., 2023)	72	75	60	63	46.79
YOLOv5-M (Yang et al., 2023)	71	74	60	62	21.18
YOLOv5-S (Shi et al., 2024)	67	69	64	62	7.14
YOLOv5-X (Yang et al., 2023)	75	81	57	62	87.45
YOLOX-L (Xu et al., 2023)	78	79	69	71	54.17
YOLOX-M (Liu et al., 2024)	78	80	73	75	25.30
YOLOX-S (Zheng et al., 2023)	77	76	78	77	8.95
YOLOX-X (Guijin et al., 2024)	62	76	78	77	99.02
EfficientDet-D0 (Chen et al., 2024)	73	78	69	70	3.85
EfficientDet-D1 (Yan et al., 2025)	71	72	69	68	6.58
EfficientDet-D2 (Afif et al., 2022)	74	70	74	69	8.08
EfficientDet-D3 (Sajid et al., 2021)	71	64	73	67	11.98
YOLOv3 (Ma et al., 2020)	80	84	80	81	61.69
YOLOv4 (Shi et al., 2021)	80	81	81	81	64.10
YOLOv7 (Wei et al., 2023)	76	78	77	77	71.02
SDBN-HCWO (Proposed)	87	91	88	89	17.2

Table 7
The performance comparison of SDBN-HCWO with integrating DenseNet121, DSC (Depth-wise Separable Convolution), and H-Swish activation function.

Detector	DenseNet121	DSC	H-Swish	AP (%)	Precision (%)	Recall (%)	F1-score (%)
SDBN-HCWO	×	×	×	83.72	83.13	80.86	81.46
	✓	×	×	86.40	85.62	82.40	83.97
	×	✓	×	84.36	84.53	78.48	80.52
	×	×	✓	85.19	85.22	78.19	80.68
	✓	✓	×	86.83	87.33	84.26	85.49
	✓	×	✓	87.02	85.84	84.29	84.92
	×	✓	✓	85.45	85.95	83.51	84.59
	✓	✓	✓	87.43	90.71	87.61	88.96

any additional components, achieved an average precision of 83.72 %, with a Precision of 83.13 %, Recall of 80.86 %, and an F1-score of 81.46 %. When DenseNet121 was added by itself, there was a significant improvement in performance, with the AP increasing to 86.40 % and the F1-score rising to 83.97 %. Adding either DSC or H-Swish individually also led to modest gains, resulting in AP values of 84.36 % and 85.19 %, respectively. The most substantial improvements were observed when both DenseNet121 and DSC were combined, which achieved an AP of 86.83 %, and when DenseNet121 and H-Swish were integrated, resulting in an AP of 87.02 %. However, the best performance overall was achieved by incorporating all three components simultaneously, yielding an AP of 87.43 %, a Precision of 90.71 %, a Recall of 87.61 %, and an F1-score of 88.96 %. These results indicate that the synergistic integration of DenseNet121, DSC, and H-Swish significantly enhances damage detection accuracy, making the model more effective at identifying structural damage.

4. Discussion and future implication

This research utilized satellite imagery to improve the accuracy and representation of the building detection process. In urban environments, precise building representation is essential for facilitating rural development. The integration of machine learning

techniques significantly enhances the efficiency of building detection from satellite images. Furthermore, deep learning approaches are particularly effective in increasing detection accuracy while reducing training time. The development of methodologies based on machine learning and deep learning for building detection aims to optimize performance and minimize the false positive rate.

To achieve robust building detection, this research introduced the SDBN-HCWO method. By integrating Discrete Latent Deep Reinforcement Learning with a discrete latent function, the method enhances detection accuracy. Additionally, the bubble-net mechanism decreases the false positive rate by generating distinct, helix-shaped bubbles that connect detected objects. The Secant Object Detection model is then applied to identify buildings based on the outputs of the bubble-net mechanism. Simulation results demonstrate that the SDBN-HCWO method classifies buildings with higher accuracy and efficiency, requiring less time and producing fewer false positives compared to existing approaches.

This research presents a comprehensive comparative analysis to evaluate the performance of the proposed building detection method relative to conventional approaches. A diverse set of sample images is used in the experimental analysis to assess the effectiveness of the method. The proposed SDBN-HCWO method is benchmarked against multiple performance metrics, including Peak Signal-to-Noise Ratio (PSNR), CA, CT, and FPR. The results indicate that SDBN-HCWO outperforms conventional methods, achieving an 18 % improvement in PSNR, a 34 % increase in CA, a 19 % reduction in training time, and a 58 % decrease in the false positive rate.

The performance evaluation of the SDBN-HCWO model highlights its robustness and effectiveness in detecting building damage across various disaster scenarios. The model achieves high detection accuracy, particularly for completely destroyed and non-damaged buildings, with average precision (AP) scores of 90.40 % and 92.67 %, respectively, ensuring reliable identification of these extreme cases. Detection for lightly damaged buildings is consistent, with an AP of 88.13 %. However, identifying medium-damaged structures poses challenges, as indicated by a lower AP of 78.51 %. This lower score suggests that while the model achieves the highest precision of 96.31 %, it tends to miss some cases, reflected in a lower recall rate of 79.00 %. In geographic terms, SDBN-HCWO performs exceptionally well in hurricane and flood-affected areas, with impressive AP scores of 90.24 % for the Nepal flooding and 90.22 % for Hurricane Florence. However, the model shows limitations in detecting damage from tornadoes and volcanic disasters, such as in the Joplin tornado (Recall: 81.84 %) and Guatemala volcano (F1-score: 83.64 %), likely due to the complex patterns of structural damage. When compared to existing models, SDBN-HCWO stands out by achieving the highest AP (87 %), precision (91 %), recall (88 %), and F1-score (89 %). It surpasses YOLOv3, YOLOv4, and Faster R-CNN, all while maintaining a manageable parameter size of 17.2 MB. These findings indicate that SDBN-HCWO marks a significant advancement in damage detection accuracy and efficiency, although further refinements in medium-damage detection and recall optimization could further enhance its overall performance.

While the dataset utilized in this study is relatively small, it is highly relevant to the problem at hand and yields strong detection results. However, the performance of the method diminishes when applied to larger, more complex images, indicating limitations in its generalization ability. Future work should focus on integrating advanced algorithms to address this issue. Specifically, incorporating lightweight models into the backbone or detection module could reduce CT without compromising detection accuracy. Additionally, considering multiple objective functions could help identify optimal boundaries for building detection. Future research could also explore the efficiency of the proposed model by testing it with different deep learning architectures. Furthermore, expanding the study to include larger and more diverse datasets would provide a clearer evaluation of the scalability of the approach. Lastly, the computational overhead associated with building detection could be mitigated in future studies through more efficient preprocessing techniques.

5. Conclusions

The results of the proposed SDBN-HCWO method demonstrate significant advancements in building detection by effectively integrating a bio-inspired optimization technique with deep learning. The method leverages the Secant Deep Belief Network's hidden layers to capture complex features, while the Hyperbolic Cosine Whale Optimization (HCWO) refines the detection process through specialized mechanisms such as the Hyperbolic Cosine Prey Encircling function and the Shrinking Encircle method. These novel components collectively enhance edge detection and linking precision, leading to more accurate and reliable identification of building structures.

Experimental evaluations on the Massachusetts Building dataset show that SDBN-HCWO achieves superior performance compared to traditional methods, evidenced by improved metrics including higher Correct Accuracy (CA) and Peak Signal-to-Noise Ratio (PSNR), alongside reduced Computation Time (CT) and False Positive Rate (FPR). This highlights the method's capability to balance accuracy with computational efficiency, an essential requirement for practical remote sensing applications. Despite these promising outcomes, some limitations remain, such as potential sensitivity to varying image resolutions or complex urban landscapes, which may affect edge detection robustness. Future research could explore adaptive mechanisms to address these challenges, as well as extend the framework to other domains like disaster assessment or urban planning. Additionally, incorporating real-time processing capabilities and investigating the integration with other sensor data (e.g., LiDAR) may further enhance the applicability of SDBN-HCWO. Overall, this research contributes a novel, efficient, and accurate solution for automated building detection, with strong potential for diverse remote sensing applications.

Despite the promising results achieved by the proposed SDBN-HCWO method, several limitations were observed during experimentation. One of the primary constraints lies in the relatively small size of the dataset used in this study. While the dataset is highly relevant and contributes to strong detection outcomes, the model's performance noticeably declines when applied to larger and more complex images. This suggests a limitation in the generalization capability of the current approach. Moreover, the computational time (CT) remains a challenge, particularly in scenarios involving high-resolution images or large-scale datasets. To address these limitations, future research should focus on integrating advanced algorithms that can enhance both scalability and efficiency. One potential

direction involves incorporating lightweight neural network models into the backbone or detection module, which could help reduce computational overhead without sacrificing detection accuracy. Additionally, exploring multi-objective optimization strategies may lead to more effective boundary identification, thereby improving the overall precision of building detection. Further evaluation of the model's efficiency could also be conducted by testing it with a range of alternative deep learning architectures to determine the most robust and scalable configuration. Expanding the dataset to include a broader range of urban environments and architectural styles would enhance the model's ability to generalize across different contexts. Lastly, optimizing preprocessing techniques could help minimize computational burden, making the approach more suitable for real-time or large-scale remote sensing applications.

CRediT authorship contribution statement

Md Helal Miah: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Shuanggen Jin:** Supervision, Formal analysis, Data curation, Conceptualization. **Mayin Uddin Jubaid:** Resources, Project administration, Methodology, Investigation. **Md Altab Hossin:** Validation, Investigation, Funding acquisition, Conceptualization.

Ethical Statement for Solid State Ionics

1) This material is the authors' own original work, which has not been previously published elsewhere.
2) The paper is not currently being considered for publication elsewhere.
3) The paper reflects the authors' own research and analysis in a truthful and complete manner.
4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
5) The results are appropriately placed in the context of prior and existing research.
6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

The violation of the Ethical Statement rules may result in severe consequences.

To verify originality, your article may be checked by the originality detection software iThenticate. See also http://www.elsevier.com/editors/plagdetect.

I agree with the above statements and declare that this submission follows the policies of Solid State Ionics as outlined in the Guide for Authors and in the Ethical Statement.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors are grateful to the anonymous reviewer for their valuable comments. This work is supported by the National Natural Science Foundation of China (Grant No. 41771566), the Humanities and Social Science Project Funding of the Education Ministry of China (Project number: 16YJC630081) and the technology innovative team on water resources safety and clean energy collaborative management of Henan province (Henan provincial science and technology department). The authors thank Shuanggen Jin for enlightening discussions.

Data availability

Data will be made available on request.

References

Abdollahi, A., Pradhan, B., Alamri, A.M., 2022. An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. Geocarto Int. 37, 3355–3370. https://doi.org/10.1080/10106049.2020.1856199.

Afif, M., Ayachi, R., Said, Y., Atri, M., 2022. An evaluation of EfficientDet for object detection used for indoor robots assistance navigation. J. Real Time Image Process 19, 651–661. https://doi.org/10.1007/S11554-022-01212-4/METRICS.

Argyrou, A., Agapiou, A., 2022. A review of artificial intelligence and remote sensing for archaeological research. Remote Sens. 14, 6000. https://doi.org/10.3390/RS14236000, 2022;14:6000.

Ayemowa, M.O., Ibrahim, R., Bena, Y.A., 2024. A systematic review of the literature on deep learning approaches for cross-domain recommender systems. Decision Anal. J. 13, 100518. https://doi.org/10.1016/J.DAJOUR.2024.100518.

Balyan, A.K., Ahuja, S., Lilhore, U.K., Sharma, S.K., Manoharan, P., Algarni, A.D., et al., 2022. A hybrid intrusion detection model using EGA-PSO and improved random forest method. Sensors 22, 5986. https://doi.org/10.3390/S22165986, 2022;22:5986.

Basha, S.J., Logu, K., 2024. Comparison of natural language processing algorithm with support vector machine for fake news identification to improve peak signal to noise ratio with classified accuracy. AIP Conf. Proc. 2853. https://doi.org/10.1063/5.0197642/3290153.
Bhardwaj, D., Nagabhooshanam, N., Singh, A., Selvalakshmi, B., Angadi, S., Shargunam, S., et al., 2025. Enhanced satellite imagery analysis for post-disaster building damage assessment using integrated ResNet-U-Net model. Multimed. Tool. Appl. 84, 2689–2714. https://doi.org/10.1007/S11042-024-20300-0/METRICS.
Cao, D., Xu, Y., Yang, Z., Dong, H., Li, X., 2023. An enhanced whale optimization algorithm with improved dynamic opposite learning and adaptive inertia weight strategy. Complex Intell. Syst. 9, 767–795. https://doi.org/10.1007/S40747-022-00827-1/TABLES/17.
Chen, J., Lu, W., Lou, J., 2023a. Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. Comput. Aided Civ. Infrastruct. Eng. 38, 1079–1098. https://doi.org/10.1111/MICE.12928.
Chen, S., Ogawa, Y., Zhao, C., Sekimoto, Y., 2023b. Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach. ISPRS J. Photogrammetry Remote Sens. 195, 129–152. https://doi.org/10.1016/J.ISPRSJPRS.2022.11.006.
Chen, S., Liu, M., Hu, L., Wang, M., 2024. A pedestrian detection network based on EfficientDet combined with SCConv. 2024 3rd international conference on robotics, artificial intelligence and intelligent control. RAIIC 378–383. https://doi.org/10.1109/RAIIC61787.2024.10671013.
Chou, J.S., Molla, A., 2022. Recent advances in use of bio-inspired jellyfish search algorithm for solving optimization problems. Sci. Rep. 12 (1), 1–23. https://doi.org/10.1038/s41598-022-23121-z, 2022;12.
Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., et al., 2021. Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. Neural Comput. Appl. 33, 6697–6720. https://doi.org/10.1007/S00521-020-05449-7/TABLES/2.
Deep learning for satellite image processing n.d. https://www.kaggle.com/code/bloodaxe/deep-learning-for-satellite-image-processing#Deep-learning-for-satellite-image-processing (accessed March 29, 2025).
Dhanaraj, R.K., Ramakrishnan, V., Poongodi, M., Krishnasamy, L., Hamdi, M., Kotecha, K., et al., 2021. Random forest bagging and X-Means clustered antipattern detection from SQL query log for accessing secure Mobile data. Wireless Commun. Mobile Comput. 2021, 2730246. https://doi.org/10.1155/2021/2730246.
Ding, J., Zhang, J., Zhan, Z., Tang, X., Wang, X., 2022. A precision efficient method for collapsed building detection in post-earthquake UAV images based on the improved NMS algorithm and faster R-CNN. Remote Sens. 14, 663. https://doi.org/10.3390/RS14030663, 2022;14:663.
Dixit, M., Chaurasia, K., Kumar Mishra, V., 2021. Dilated-ResUnet: a novel deep learning architecture for building extraction from medium resolution multi-spectral satellite imagery. Expert Syst. Appl. 184, 115530. https://doi.org/10.1016/J.ESWA.2021.115530.
Fang, L., Liang, X., 2023a. A novel method based on nonlinear binary grasshopper whale optimization algorithm for feature selection. J. Bionic. Eng. 20, 237–252. https://doi.org/10.1007/S42235-022-00253-6/TABLES/12.
Fang, L., Liang, X., 2023b. A novel method based on nonlinear binary grasshopper whale optimization algorithm for feature selection. J. Bionic. Eng. 20, 237–252. https://doi.org/10.1007/S42235-022-00253-6/METRICS.
Guijin, H., Ruixuan, W., WuYan, X., Jun, L., 2024. Night construction site detection based on ghost-YOLOX. Connect. Sci. 36. https://doi.org/10.1080/09540091.2024.2316015.
Günen, M.A., 2024. Fast building detection using new feature sets derived from a very high-resolution image, digital elevation and surface model. Int. J. Rem. Sens. 45, 1477–1497. https://doi.org/10.1080/01431161.2024.2313991;CTYPE:STRING:JOURNAL.
Günen, M.A., Pérez-Delgado, M.-L., Beşdok, E., Günen, M.A., Pérez-Delgado, M.-L., Beşdok, E., 2024. L0-Norm based image pansharpening by using population-based algorithms. AIMS Math. 11 (9), 32578–32628. https://doi.org/10.3934/MATH.20241561, 32578 2024.
Günen, M.A., Öztürk, K.F., Aliyazıcıoğlu, Ş., 2025. Optimizing visibility of historical structures using mWDE: insights from the Kromni Valley, Gümüşhane, Türkiye. Int. J. Electron. Govern. 10, 107–126. https://doi.org/10.26833/IJEG.1529351.
Han, Q., Yin, Q., Zheng, X., Chen, Z., 2022. Remote sensing image building detection method based on Mask R-CNN. Complex Intell. Syst. 8, 1847–1855. https://doi.org/10.1007/S40747-021-00322-Z/FIGURES/5.
Hossain, S.M.M., Deb, K., Dhar, P.K., Koshiba, T., 2021. Plant leaf disease recognition using depth-wise separable convolution-based models. Symmetry 13, 511. https://doi.org/10.3390/SYM13030511, 2021;13:511.
Huang, X., Cao, Y., Li, J., 2020. An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images. Remote Sens. Environ. 244, 111802. https://doi.org/10.1016/J.RSE.2020.111802.
Huang, J., Feng, W., Sun, Y., Wang, H., Yan, J., Deng, J., et al., 2025. Deep learning-based building change detection in off-nadir images via a pixel-wise and patch-wise fusion strategy. Trans. GIS 29, e70020. https://doi.org/10.1111/TGIS.70020.
Javed, A.R., Fahad, L.G., Farhan, A.A., Abbas, S., Srivastava, G., Parizi, R.M., et al., 2021. Automated cognitive health assessment in smart homes using machine learning. Sustain. Cities Soc. 65, 102572. https://doi.org/10.1016/J.SCS.2020.102572.
Ju, S., Chen, Z., Liao, X., Shen, Y., He, J., Su, Y., 2024. A tiny efficient U-Net with gated linear attention for medical image segmentation. In: Proceedings - 2024 IEEE International Conference on Bioinformatics and Biomedicine. BIBM, pp. 3367–3371. https://doi.org/10.1109/BIBM62325.2024.10822137.
Khankeshizadeh, E., Mohammadzadeh, A., Arefi, H., Mohsenifar, A., Pirasteh, S., Fan, E., et al., 2024. A novel weighted ensemble transferred U-Net based model (WETUM) for postearthquake building damage assessment from UAV data: a comparison of deep learning-and machine learning-based approaches. IEEE Trans. Geosci. Rem. Sens. 62, 1–17. https://doi.org/10.1109/TGRS.2024.3354737.
Kononchuk, R., Cai, J., Ellis, F., Thevamaran, R., Kottos, T., 2022. Exceptional-point-based accelerometers with enhanced signal-to-noise ratio. Nature 607 (607), 697–702. https://doi.org/10.1038/s41586-022-04904-w, 7920 2022.
Langerman, D., Johnson, A., Buettner, K., George, A.D., 2020. Beyond floating-point ops: CNN performance prediction with critical datapath length. In: 2020 IEEE High Performance Extreme Computing Conference. HPEC. https://doi.org/10.1109/HPEC43674.2020.9286182.
Li, Y., Hu, W., Dong, H., Zhang, X., 2019. Building damage detection from post-event aerial imagery using single shot Multibox detector. Appl. Sci. 9, 1128. https://doi.org/10.3390/APP9061128, 2019;9:1128.
Li, J., Wang, J., Lin, F., Wu, W., Chen, Z.M., Asghar Heidari, A., et al., 2024. DSEUNet: a lightweight UNet for dynamic space grouping enhancement for skin lesion segmentation. Expert Syst. Appl. 255, 124544. https://doi.org/10.1016/J.ESWA.2024.124544.
Liu, W., Zhang, S., Zhou, L., Luo, N., Xu, M., 2024. YMMNet: a more accurate and lightweight detector of illegal buildings for smart cities. IEEE Trans. Consum. Electron. https://doi.org/10.1109/TCE.2024.3415673.
Luis, C., González, M., Montoya, G.A., Lozano Garzón, C., 2025. Toward reliable post-disaster assessment: advancing building damage detection using you only look once convolutional neural network and satellite imagery. Mathematics 13, 1041. https://doi.org/10.3390/MATH13071041, 2025;13:1041.
Luo, L., Li, P., Yan, X., 2021. Deep learning-based building extraction from remote sensing images: a comprehensive review. Energies 14, 7982. https://doi.org/10.3390/EN14237982, 2021;14:7982.
Ma, H., Liu, Y., Ren, Y., Yu, J., 2020. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3. Remote Sens. 12, 44. https://doi.org/10.3390/RS12010044, 2019;12:44.
Massachusetts buildings dataset. https://www.kaggle.com/datasets/balraj98/massachusetts-buildings-dataset. (Accessed 21 February 2025).
Mozafari, M., Moattar, M.H., 2025. A hybrid fuzzy deep belief network extreme learning machine framework with hyperbolic secant activation function for robust semi-supervised sentiment classification. Appl. AI Lett. 6, e102. https://doi.org/10.1002/AIL2.102.
Pin Tan, D.K., He, J., Li, Y., Bayesteh, A., Chen, Y., Zhu, P., et al., 2021. Integrated sensing and communication in 6G: motivations, use cases, requirements, challenges and future directions. In: 2021 1st IEEE International Online Symposium on Joint Communications and Sensing, JC and S 2021. https://doi.org/10.1109/JCS52304.2021.9376324.
PushpaRani, K., Roja, G., Anusha, R., Dastagiraiah, C., Srilatha, B., Manjusha, B., 2024. Geological information extraction from satellite imagery using deep learning. In: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7. https://doi.org/10.1109/ICCCNT61001.2024.10724802.
Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R.R., 2021. FloodNet: a high resolution aerial imagery dataset for post flood scene understanding. IEEE Access 9, 89644–89654. https://doi.org/10.1109/ACCESS.2021.3090981.

Rai, A.K., Mandal, N., Singh, K.K., Izonin, I., 2023. Satellite image classification using a hybrid manta ray foraging optimization neural network. Big Data Mining Anal. 6, 44–54. https://doi.org/10.26599/BDMA.2022.9020027.

Sajid, F., Javed, A.R., Basharat, A., Kryvinska, N., Afzal, A., Rizwan, M., 2021. An efficient deep learning framework for distracted driver detection. IEEE Access 9, 169270–169280. https://doi.org/10.1109/ACCESS.2021.3138137.

Sanhudo, L., Calvetti, D., Martins, J.P., Ramos, N.M.M., Mêda, P., Gonçalves, M.C., et al., 2021. Activity classification using accelerometers and machine learning for complex construction worker activities. J. Build. Eng. 35, 102001. https://doi.org/10.1016/J.JOBE.2020.102001.

Senussi, M.F., Kang, H.S., 2024. Occlusion removal in light-field images using CSPDarknet53 and bidirectional feature pyramid network: a multi-scale fusion-based approach. Appl. Sci. 14, 9332. https://doi.org/10.3390/APP14209332, 2024;14:9332.

Sharma, D., Singhai, J., 2021. An unsupervised framework to extract the diverse building from the satellite images using grab-cut method. Earth Sci. Inform 14, 777–795. https://doi.org/10.1007/S12145-021-00569-7/METRICS.

Shi, L., Zhang, F., Xia, J., Xie, J., Zhang, Z., Du, Z., et al., 2021. Identifying damaged buildings in aerial images using the object detection method. Remote Sens. 13, 4213. https://doi.org/10.3390/RS13214213, 2021;13:4213.

Shi, Y., Luo, Q., Zhou, M., Guo, W., Li, J., Li, S., et al., 2024. S-YOLOv5: a lightweight model for detecting objects thrown from tall buildings in communities. Information 15, 188. https://doi.org/10.3390/INFO15040188, 2024;15:188.

Singh, N.J., Nongmeikapam, K., 2022. Semantic segmentation of satellite images using deep-unet. Arabian J. Sci. Eng. 48, 1193–1205. https://doi.org/10.1007/S13369-022-06734-4.

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., et al., 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery.

Sun, X., Yin, D., Qin, F., Yu, H., Lu, W., Yao, F., et al., 2023. Revealing influencing factors on global waste distribution via deep-learning based dumpsite detection from satellite imagery. Nat. Commun. 14 (1), 1–13. https://doi.org/10.1038/s41467-023-37136-1, 2023;14.

Tanim, A.H., McRae, C.B., Tavakol-davani, H., Goharian, E., 2022. Flood detection in urban areas using satellite imagery and machine learning. Water 14, 1140. https://doi.org/10.3390/W14071140, 2022;14:1140.

Uddin, S., Haque, I., Lu, H., Moni, M.A., Gide, E., 2022. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci. Rep. 12 (12), 1–11. https://doi.org/10.1038/s41598-022-10358-x, 1 2022.

Walter, I., Tanasković, M., Stanković, M., 2023. IR building analysis with extraction of elements using image segmentation and RetinaNet. Buildings 13, 109. https://doi.org/10.3390/BUILDINGS13010109, 2022;13:109.

Wei, G., Wan, F., Zhou, W., Xu, C., Ye, Z., Liu, W., et al., 2023. BFD-YOLO: a YOLOv7-Based detection method for building façade defects. Electronics 12, 3612. https://doi.org/10.3390/ELECTRONICS12173612, 2023;12:3612.

Wu, Y., Zou, B., Cao, Y., Wu, Y., Zou, B., Cao, Y., 2024. Current status and challenges and future trends of deep learning-based intrusion detection models. J. Imaging 10. https://doi.org/10.3390/JIMAGING10100254. Page 254 2024;10:254.

Xie, Y., Zhan, N., Zhu, J., Xu, B., Chen, H., Xie, Y., Zhan, N., Zhu, J., Xu, B., Chen, H., Mao, W., Luo, X., Hu, Y., et al., 2024. Landslide extraction from aerial imagery considering context association characteristics. Int. J. Appl. Earth Obs. Geoinf. 131, 103950.

Xu, Y., Zhang, J., 2024. A hybrid nonlinear whale optimization algorithm with sine cosine for global optimization. Biomimetics 9, 602. https://doi.org/10.3390/BIOMIMETICS9100602, 2024;9:602.

Xu, Y., Jiang, Y., Li, L., Wang, Y., Li, R., Cai, B., et al., 2023. An improved YOLOX autonomous UAV-based system for monitoring photovoltaic panel failures. In: 2023 5th International Conference on Frontiers Technology of Information and Computer, ICFTIC 2023, pp. 330–350. https://doi.org/10.1109/ICFTIC59930.2023.10456038.

Yan, Q., Zhang, H., Zhao, H., 2025. A new indoor occupancy detection model by integrating the efficient multi-scale attention mechanism into the EfficientDet model. Commun. Comput. Inform. Sci. 2181, 132–145. https://doi.org/10.1007/978-981-97-7001-4_10. CCIS.

Yang, X., Xie, Y., Yang, S., Liang, P., He, Y., Yang, J., et al., 2023. Research on application of object detection based on yolov5 in construction site. 2023 15th international conference on advanced computational intelligence. ICACI 2023. https://doi.org/10.1109/ICACI58115.2023.10146151.

Zeng, Y., Guo, Y., Li, J., 2022. Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning. Neural Comput. Appl. 34, 2691–2706. https://doi.org/10.1007/S00521-021-06027-1/METRICS.

Zhang, Y., 1999. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. ISPRS J. Photogrammetry Remote Sens. 54, 50–60. https://doi.org/10.1016/S0924-2716(98)00027-6.

Zheng, X., Fang, S., Chen, H., Peng, L., Ye, Z., 2023. Internal detection of ground-penetrating radar images using YOLOX-s with modified backbone. Electronics 12, 3520. https://doi.org/10.3390/ELECTRONICS12163520, 2023;12:3520.