Vietnam Journal of Earth Sciences, 48(4), 561−586, https://doi.org/10.15625/2615-9783/24801

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

UAV and deep-learning-based coastal debris detection for supporting proposed 4.0 management strategies in Vietnam

Thi Ngan Do[1,2], Kinh Bac Dang[1,*], Kim Chi Vu[3], Tuan Linh Giang[1,3], Thi Phuong Nga Pham[1], Thi Yen Nguyen[1], Thi Dieu Linh Nguyen[1], Van Trong Giang[3], Minh Hieu Nguyen[1]

[1]*VNU-University of Sciences, Vietnam National University, Hanoi, Vietnam*
[2]*Cau Giay Secondary School, Hanoi, Vietnam*
[3]*VNU-Institute of Vietnamese Studies and Development Science (VNU-IVIDES), Vietnam National University, Hanoi, Vietnam*

Received 05 April 2026; Received in revised form 18 May 2026; Accepted 15 June 2026

ABSTRACT

Coastal debris directly impacts people's health and the environment, so assessing it is crucial to industrialization, modernization, and economic development while conserving the environment for stable, long-term prosperity. Using drones to photograph debris on different coasts in Vietnam from 2023 to 2025, this study aims to develop deep learning (DL) models to assess coastal debris distribution. The UNet and PSPN using a ResNet34 backbone achieved better debris distribution identification with an input size of 64 × 64 pixels. The assessment indicated that the debris was primarily concentrated in littoral areas and embankments, and that it fluctuated with seasonal variations and collection efforts. The majority of debris was found to have been generated by fishing vessels, river systems, and marine trading activities, as evidenced by local surveys. Coastal debris, including nylon bags, was highly visible throughout the site. Current debris management is inefficient because manual periodic collection methods lack sufficient disposal solutions, and local groups struggle to collect trash. The research found that both local authorities and residents were deficient in exchanging environmental information effectively due to poor ecological awareness among the public. Therefore, the proposed framework should be regarded as a management recommendation rather than a field-tested operational system. This proposal can help increase environmental monitoring and engagement, enabling local authorities to work more effectively with their communities to implement effective debris management practices.

Keywords: Debris, artificial intelligence, coast, Nam Dinh, Vietnam.

1. Introduction

Currently, the coastal debris has emerged as a global concern (Zia et al., 2022). Coastal debris directly impacts the environment and the livelihoods of individuals (Molnár et al., 2024; Wahid and Mutaqin, 2024). The management of this debris type is an urgent issue in the context of promoting industrialization, modernization, and economic development, while ensuring environmental protection to guarantee stable and sustainable development (Sakti et al., 2021; Suhail et al., 2023).

Debris pollution is becoming an urgent and concerning issue along Vietnam's 3,200-kilometer coastline (Thi Nham, 2024). In terms of coastal debris pollution, Vietnam ranks fourth globally. People in some provinces recorded over 30,000 tons of

*Corresponding author, Email: dangkinhbac@hus.edu.vn

coastal solid debris per year (Tsai et al., 2020). Coastal environments and human health are severely damaged by the unchecked accumulation of debris. Daily debris entering the ocean accumulates from various sources, including river runoff and debris from fisheries and tourism in coastal areas (Ayub and Ren, 2024). The debris problem continues to worsen with the rising human population and the growth of economic activity (Giuliani et al., 2019). It proves that multiple scientific research groups working under the international community methodically studied and assessed this issue.

Kubiszewski et al. (2016) noted that government organizations must maintain ongoing relationships with NGOs, businesses, and communities to advance effective, sustainable solutions. The implementation of 4.0 technology-based debris management systems and monitoring devices proves effective for solving this problem (Iqbal et al., 2022). Deep learning (DL) systems enable coastal management teams to detect environmental problems (Luo et al., 2022; Nguyen et al., 2020) associated with pollution and coastal damage (Dang et al., 2022c) and environmental changes (Dang et al., 2020b; Matching et al., 2026) through advanced models for prediction and image processing systems. DL technology aids environmental management by enabling controllers to outsource data acquisition while performing analytics with high accuracy throughout the interpretation process. The process enables monitoring changes in nature and making timely, appropriate decisions (Pucino et al., 2022; Seale et al., 2022; Yasir et al., 2024). Effective future protection of coastal environments and community welfare demands fresh scientific inquiry, which needs rapid research investigations (Berto et al., 2025; Merlino et al., 2021). Wildlife encounters problems due to pollution through two mechanisms: first, they consume pollutants, and second, they disrupt their habitats. Therefore, tracking debris along coasts is an essential procedure that protects coastal environments (Choi et al., 2023). Debris monitoring provides scientists with insights into its ecological effects and enables rapid cleanup initiation and enhanced protection of water quality (Nunkhaw and Miyamoto, 2024). Additionally, observing debris serves two crucial functions: protecting the environment, promoting public involvement, and supporting sustainable tourism activities.

Although uncrewed aerial vehicle (UAV) imagery, DL models, and citizen-science approaches have increasingly been used for marine litter and coastal debris monitoring, previous studies have mainly emphasized detection accuracy and image-based classification (Yasir et al., 2026). Less attention has been paid to linking debris detection results with coastal management needs and to interpreting debris accumulation from a spatial-geomorphological perspective. This is an important gap because coastal debris is not randomly distributed; it is transported and trapped by shoreline morphology, tidal flats, embankments, waves, estuarine processes, and longshore currents. In Vietnam, UAV-based DL studies that connect debris detection, spatial interpretation, and management-oriented recommendations remain limited.

This study aims to develop and evaluate deep learning (DL) models for detecting coastal debris in UAV images and to discuss how the resulting debris maps can support a proposed 4.0 technology-based management framework for coastal debris monitoring in Vietnam. The main contributions are: (i) developing a UAV-derived dataset for coastal debris detection; (ii) comparing DL architectures and input sizes; (iii) assessing debris changes after cleaning and across seasons; and (iv) proposing a 4.0 technology-

based management framework. This advanced technology combination promotes sustainable practices, community awareness, and informed decision-making, making it a powerful tool in the ongoing global initiative to combat pollution and protect coastal ecosystems. Additionally, this study could evaluate the existing circumstances by conducting data acquisition from agencies and government and non-governmental organizations conducting direct surveys of river flows and coasts; evaluate the extent of pollution and pinpoint the primary causes of this issue through data analysis and investigation; and identify the specific causes of coastal debris contamination in a particular region in Vietnam by employing research methodologies such as surveys, interviews, and data analysis. Many aspects contribute to high levels of pollution from debris, including the disposal practices of local communities and their exorbitant consumption patterns, as well as insufficient debris treatment systems.

2. Methodology

2.1. Research area

Various coastal regions in Vietnam were selected to collect data for training DL models, including Quang Ninh, Nam Dinh, Da Nang, and Quang Nam (Fig. 1). However, the chosen research area for testing the final model is the coast along Giao Hai district in Nam Dinh province (Nam et al., 2022). Nam Dinh province has a 72-kilometer coastline, natural beauty, and a favorable geographical location, all of which contribute to the city's economic growth and ecotourism industry, which annually draws millions of visitors. In recent years, however, debris along the coast has caused substantial environmental pollution, negatively impacting local development (Maddalene et al., 2021).

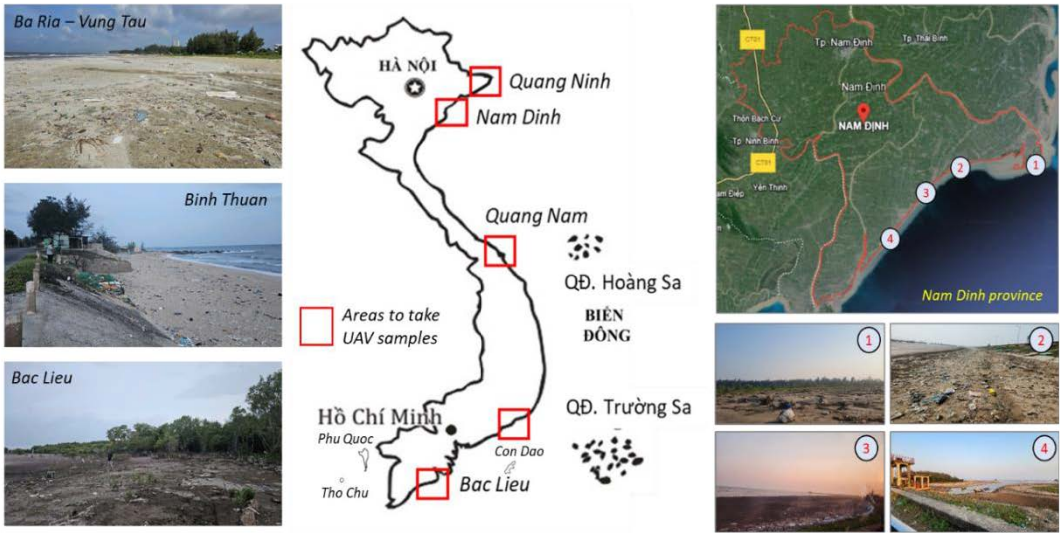

Figure 1. Location of the chosen research area along the Giao Hai coast in Nam Dinh province

In Vietnam, the Youth Union volunteers to carry out the coastal cleaning campaign have been organized by the People's Committee of coastal provinces and the Vietnam Union of Science and Technology Associations. The main types of debris include bags, bottles, glass jars, rubber tires, and foam mattresses, which are scattered across coastal environments, including beaches and dunes. This situation has not been addressed for a long time, not only causing serious environmental pollution but also eroding coastal beauty. Coastal communes, particularly in Nam Dinh province, have

implemented many coordinated programs to clean the coastal area, educate fishermen on the conscious use of glass products, and encourage voluntary collection and classification of inorganic and organic debris (Tran et al., 2020). Many programs, such as: Youth union members cleaning the coastal area, building new rural communes, and building model new rural communes. All have integrated and guided people to protect the marine environment. However, due to a lack of regular maintenance, coastal debris continues to accumulate.

According to many fishermen, this type of debris is largely discarded by boat fishing activities, aquaculture households, and shipping vessels, and is washed ashore by waves (WWF, 2023). For a long time, no one collected it, forming debris dumps along the coast. For many years, the debris along the Giao Hai coast has not been addressed. All collection and treatment activities are mainly carried out by local people about three times per year, using self-contributed funds and on-site debris burning. Locals attribute a portion of the cause to foreign debris that large estuaries carry into the ocean. Despite a sign prohibiting trash disposal in Giao Hai district, impromptu landfills continue to form, containing a variety of materials, including foam boxes, bottles, bags, and even clam shells. Dead animals and snail shells are also dumped onto the shore, causing a foul odor (Nam et al., 2022). Therefore, a study proposing the application of DL technology could be a better solution for local managers to control the increase in debris along the coast.

2.2. Research framework

The DL model development and implementation for coastal debris management is shown in Fig. 2. Accordingly, this study employs DL models, geographic information systems (GIS), and multi-temporal remote-sensing technology, along with direct interviews, as its research methods. After synthesizing the data, the project developed criteria for detecting coastal debris in field and remote-sensing images. Combined with data collected by UAVs and interviews conducted during six field trips from 2021 to 2025 along different coasts of Vietnam, the project has determined the current status of the distribution, management, collection, and treatment of coastal debris. At the same time, the project identifies human and natural causes of coastal debris along the coasts of Nam Dinh province. It proposes short- and long-term solutions, supported by AI technology, for the management and reduction of coastal debris.

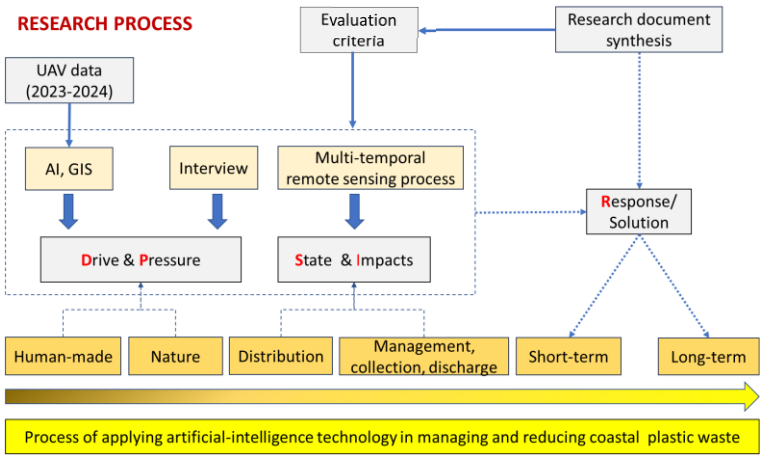

Figure 2. Framework to apply artificial-intelligence technology in managing and reducing coastal debris

2.3. Collecting and processing data

The research process through UAV data implements three essential steps. Laboratory analysis and direct site examination form the basis for debris classification. Data collection using UAV images was performed at six different times (four pre-collection observations, together with post-collection observations in 2021, 2022, 2023, and 2024; and two testing observations in 2024 and 2025) at selected ground truth points (GTP). The data collection region was along different coasts in Vietnam (Fig. 1). In total, 37 UAV images were collected. The training images were used to define visual debris characteristics, generate image-masks pairs, and train the semantic segmentation models. The testing images were independent UAV scenes used to evaluate model robustness and reduce the risk of overfitting. The Nam Dinh UAV images were used as application data to map the spatial distribution and temporal variation of coastal debris in Giao Hai commune before and after cleanup activities and across different seasons. A summary of the UAV campaigns, including acquisition time, location, number of images, survey objective, and role in the modeling workflow, is provided in Table 1.

Table 1. Summary of UAV image datasets used for model development

Cam-paign	Time	Location	No. of images	Survey objective	Role in modeling workflow
C1	6/2021	- Hoi An coast, Quang Nam - Nha Trang coast, Khanh Hoa - Coasts in Bac Lieu	20	Initial observation of coastal debris types and visual characteristics	Training and validating the dataset
C2	7/2022	Mong Cai coasts in Quang Ninh	3	Collection of UAV samples under different coastal settings	
C3	3/2023	Giao Hai coast, Nam Dinh	2	Observation before and after local debris-cleaning activities in the dry season	Training, validating, and application dataset
C4	9/2023	Different coasts in Nam Dinh	4	Observation of debris reaccumulation in the rainy season	
C5	3/2024	Different coasts in Nam Dinh	4	Observation at the end of the dry season and comparison with 2023 data	Testing/independent validation dataset
C6	4/2025	Different coasts in Nam Dinh	4	Independent testing of model transferability to new UAV images	

While the width of each image was set at 100m from shoreline to inland areas, its length depends on the distribution of coastal debris. UAV images were acquired using a DJI Mini 3 platform at a nominal flight altitude of approximately 20 m above ground level. The DJI Mini 3 camera uses a 1/1.3-inch CMOS sensor with 48 MP effective pixels and a maximum image size of 8064 × 6048 pixels. The camera has an 80° field of view, a 24 mm equivalent focal length, and an f/1.7 aperture. The ground sampling distance (GSD) was estimated from the flight altitude and camera geometry. Using the camera field of view and image dimensions, the approximate GSD can be expressed as:

$$GSD \approx \frac{\left[2H \times \tan\left(\frac{FOV}{2}\right)\right]}{N_d} = \frac{\left[2\times 20 \times \tan\left(\frac{80}{2}\right)\right]}{\sqrt{(8064^2 + 6048^2}} = 0.0033 \text{ m/pixel} \quad (1)$$

where H is the flight altitude above ground, FOV is the camera field of view, and N_d is the number of pixels along the corresponding image dimension. At a flight altitude of approximately 20 m, the theoretical GSD is on the order of several millimeters per pixel. After ortho-mosaic generation, geometric correction, and resampling, the UAV images

used for debris detection were standardized to a spatial resolution of approximately 0.005 m. This standardized resolution was adopted to maintain consistency among UAV datasets collected during different field campaigns.

These samples serve as the foundation for additional research and model training. The UAV dataset was not derived from a single coastal scene. It included UAV images collected from several coastal provinces of Vietnam, including Quang Nam, Khanh Hoa, Bac Lieu, Quang Ninh, and Nam Dinh, representing different shoreline settings, debris backgrounds, and observation periods. Therefore, the model was exposed to a degree of spatial and environmental variability typical of Vietnamese coastal conditions. However, the UAV acquisition configuration was intentionally standardized across field campaigns. The flight altitude and image-acquisition settings were controlled to produce very-high-resolution imagery with an approximate ground sampling distance of 0.005 m. This standardization was used to reduce variability caused by differences in sensor configuration, flight altitude, and image resolution, thereby allowing the model comparison to focus on architecture, input size, and debris characteristics rather than acquisition inconsistency.

The data were standardized and geometrically corrected using specialized software to generate point clouds, build digital elevation models, and produce panoramic images of the entire area where large amounts of debris are concentrated in the research area. This separation of UAV datasets ensured that the model development process was not based on a single coastal condition. The training data provided diverse examples of debris appearance across different geomorphological and environmental settings, whereas the independent test data were used to assess model transferability to new UAV scenes.

Among the UAV datasets summarized in Table 1, the Giao Hai images were used as the application dataset for evaluating the ability of the selected DL models to monitor coastal debris dynamics in Nam Dinh province (Figs. 3 and 4). UAV data were collected at four observation times: March 2023 before debris collection, March 2023 after debris collection, September 2023 during the rainy season, and March 2024 and April 2025 in the dry season. These images enabled the study to assess both the immediate effects of community-based cleanup activities and the subsequent seasonal reaccumulation of debris along the coast.

Advanced DL techniques are employed in the second phase to train computers to identify indicators of debris from a diverse array of sources. To achieve this, the training process must incorporate both conventional and innovative designs. The models are trained on samples generated in the initial stage to ensure they can accurately identify debris across a variety of contexts and circumstances (Fig. 5). These varieties undergo rigorous testing and verification to guarantee their reliability. It involves using new UAV image data to verify the models, thereby preventing issues such as underfitting or overfitting. The meticulous testing conducted after the model's construction is intended to produce a model that is both precise and robust, capable of accurately detecting coastal debris.

Lastly, the final stage is to apply the acquired models to locate and map debris across various coastal regions. The evaluated models are used to analyze UAV images and generate detailed maps that illustrate the location and quantity of debris across various regions (Fig. 5). Environmental monitoring and regulatory projects can benefit significantly from these specific maps. Such models support lawmakers and scientific researchers in creating better ways to control debris. The developed plans function to locate debris hotspots and report the situation while offering assistance to needy locations. The authors digitized debris data into vector zones for analysis during the processing phase. The raster data was generated by grouping these

locations. As "input masks," raster data were employed for DL models, while UAV data were employed as "input images". The ground-truth masks were produced through manual interpretation of UAV images. Coastal debris objects were first visually identified based on their shape, color, texture, and field observations. The identified debris areas were manually digitized as polygon features in the ArcGIS environment and then converted into raster masks with the same spatial extent and resolution as the corresponding UAV sub-images. Each mask was encoded as a binary layer, with debris pixels assigned a value of 1 and all other pixels assigned a value of 0. Trained researchers with experience in coastal remote sensing and field-based debris observation conducted the annotation. The annotated masks were checked by senior experts using UAV images and field photographs to reduce misclassification. The same annotation criteria were applied consistently across all UAV images, sites, and observation periods.

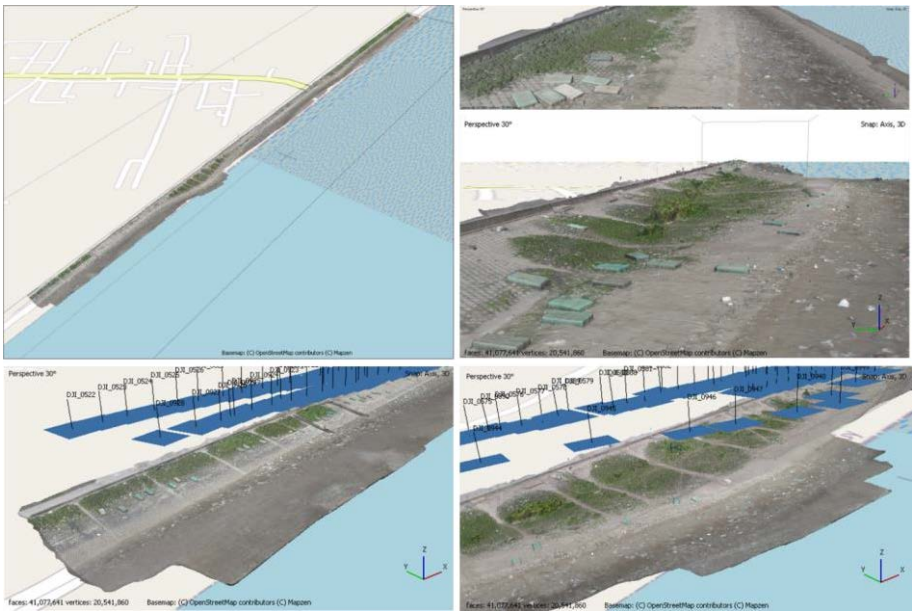

Figure 3. An example of merged images and DEM obtained from UAV data

Figure 4. Observation of coastal debris on UAV data

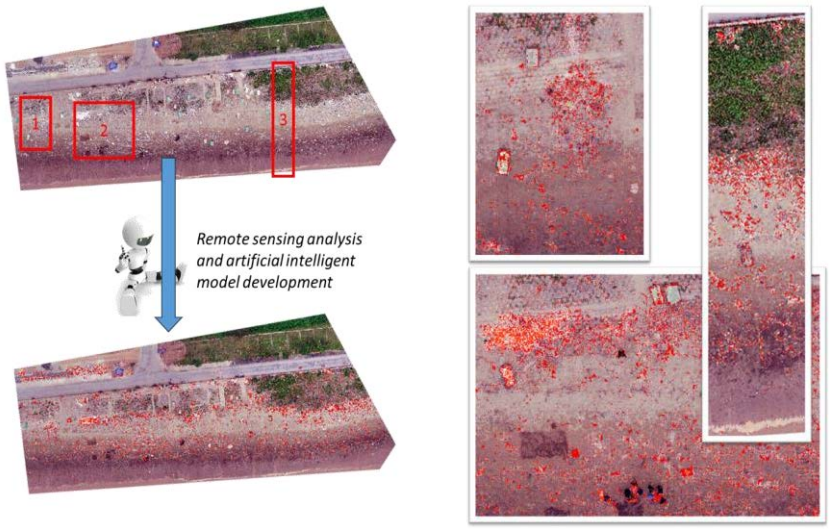

Figure 5. An example of detected coastal debris based on AI models

Data augmentation was not applied in the present experiment. This decision was made to evaluate model performance directly on the original UAV-derived sub-images and to preserve the real spectral, textural, shadow, moisture, and background characteristics of the field-acquired imagery. The absence of augmentation is acknowledged as a methodological limitation, as UAV-based debris detection can be affected by illumination, shadows, tidal conditions, moisture, surface texture, camera angle, and scale variation. However, the dataset in this study already contained a relatively large number of input sub-images generated from multiple UAVs and in different coastal cases. They can vary the brightness and contrast adjustment for the inputs. Applying augmentation at this stage would have changed the input data distribution and required retraining all model configurations, making the revised results no longer directly comparable with the original experimental design. The models were trained only on UAV-derived sub-images and the corresponding binary masks generated from the original UAV images. The validation and testing subsets were kept unchanged for model evaluation.

In this study, the term "coastal debris" refers to all visible human-made waste materials identified on UAV images. “Plastic debris, "bags", and "foam debris" are treated as specific debris types within the broader coastal debris class and were all assigned to the debris class when they appeared as abandoned or accumulated waste. In contrast, "non-debris" refers to natural surfaces and non-waste objects, including sand, soil, vegetation, water, shells, rocks, embankments, boats, shadows, and fishing or aquaculture materials that were still in active use.

Although the dataset was expanded into 32,410 image-mask sub-images, these sub-images do not represent fully independent field samples. Multiple sub-images were derived from the same original UAV scenes and may therefore share similar illumination conditions, surface textures, background characteristics, and geomorphological settings. In this study, the original UAV images were divided into non-overlapping sub-images at three input sizes: 64 × 64, 128 × 128, and 256 × 256 pixels. Small images enhance model capabilities for detecting smaller debris fragments, and large images enable models to recognize larger

clumps with their associated characteristics. For each sub-image, a corresponding binary mask was generated from manually digitized debris polygons. The generated sub-images and masks were then divided into training, validation, and testing datasets with a ratio of 70:20:10, respectively, as summarized in Table 2. DL models could process UAV images with mask data by carefully preparing samples, as demonstrated in Section 2.3. Through image processing, the computer system identifies debris and collects relevant information about the target area (Fig. 5). To reduce potential over-interpretation, model selection was based not only on accuracy, Dice, precision, recall, and F1-score, but also on validation-loss behavior, convergence stability, visual prediction quality, and model application to UAV scenes outside the main training workflow.

Table 2. Summary of UAV-derived dataset for semantic segmentation

Input size	No. of sub-images/masks	Training sub-images	Validation sub-images	Testing sub-images	Note
64 × 64	18,520	12,964	3,704	1,852	Used to detect small and scattered debris
128 × 128	9,260	6,482	1,852	926	Used to detect medium-sized debris patches
256 × 256	4,630	3,241	926	463	Used to detect larger debris accumulations
Total	**32,410**	**22,687**	**6,482**	**3,241**	

2.4. Architecture of deep-learning models

2.4.1. Deep-learning architecture components

DL models are often constructed using different architectures, such as Convolutional Neural Networks (CNNs), BiSeNet (Tsai and Tseng, 2023), and Dexinet (Sivapriya and Suresh, 2023). These architectures consist of six layers: INPUT, CONV, Batch Normalization, Pooling, Concatenate, and Dropout (Dang et al., 2022b). Before encoding, the training model assesses the raw pixel values of all sub-images in the INPUT. After processing, CONV filters can alter the incoming data. Filters modify the input pixels. Batch normalization alters the data scales. This layer reduces the need for "internal covariate shift" training or the volatility of initial values. Minimizes input loss during model training (Dang et al., 2020a). The batch mean and variance determine the standard deviation of each input layer.

$$a_i = \rho\widehat{b_i} + \omega \quad (2)$$

The values of ω and ρ are modified throughout the training process, whereas the mean and variance formulas of the mini-batch B ({b1…bn}) determine a_i. Activation algorithms for POOL layers: Perform 2×2 spatial matrix downsampling throughout the encoding process.

$$\theta_B \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i \quad (3)$$

$$\sigma_B^2 = \frac{1}{n}\sum_{i=1}^{n}(\gamma_i - \theta_B)^2 \quad (4)$$

$$\widehat{\gamma_i} \leftarrow \frac{p_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (5)$$

During the decoding process, convolution matrices, or "ConvTrans", play a role in up-sampling smaller matrices and in combining expanded route data and contraction. The larger route was simplified, and layer data was included in the contracting process before delivery. DROPOUT layers identify and deactivate redundant neurons rather than program them (Pham et al., 2024). It prevents generalization issues and under- or over-fitting (Jafarzadeh et al., 2021). Throughout the training process in this study, all DL models exhibited layer inputs with approximately 1.0 dropout. All the above

layers exist in all architectures, but they are arranged differently and have different numbers. The same training, validation, and test partitions were used for all model architectures to ensure consistent comparisons among UNet, DeepLab-v3, and Pyramid Scene Parsing Network (PSPN) models.

2.4.2. Design of UNet models

The UNet (Fig. 6) architecture is frequently employed for medical image segmentation (Dang et al., 2022a). Consequently, the term "UNet" is derived from the letter "U" and its U-shaped design. The UNet architecture consists of an encoder, a bottleneck, a decoder, skip Connections, and an output layer (Stoian et al., 2019). An encoder or contraction route is located on the U-shaped apex. The context of the input image and the features of the debris are extracted in this route. Convolutional layers and pooling enhance feature channels while reducing the input spatial dimensions. Data is processed by 3×3 convolutional layers in the contracting path. The most significant background features in UAV images, such as debris traces, were aggregated by a bottleneck layer at the base of the U-shaped structure (Fu et al., 2021; Zhang et al., 2018). This layer preserves geographic information while simultaneously reducing the dimensionality of the representation.

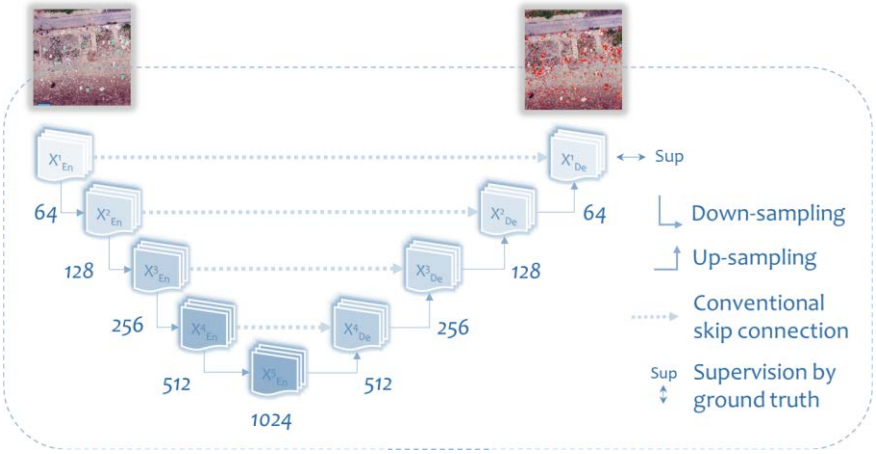

Figure 6. UNet architecture for debris detection using UAV data

By upscaling the image, this transformation is undone. This innovative filter count modification captures the context and details of critical avalanche traces (Diakogiannis et al., 2020). Deconvolution or up-sampling layers, also known as transposed convolutional layers, enhance spatial resolution and reduce the number of channels. The key to this concept is modulating the number of filters from 64 to 1024 in the early stages of the contracting route, then back to 64 in the final phase. During up-sampling, skip connections maintain the fine-grained characteristics and geography. Data exchange between encoders and decoders is simplified. The final layer of the UNet employs a 1 × 1 convolutional layer to convert the learned features to the desired output channels. The probability of debris or non-debris is indicated by the segmentation result channels.

2.4.3. Design of DeepLab-v3 models

Another architecture that improves semantic segmentation is DeepLab-v3. This architecture emerged from years of development in computer vision research by Google research teams (Chen et al., 2018a). Accordingly, the classification using this

architecture achieves high accuracy and efficiency due to numerous new features that have been included (Fig. 7). Atrous convolutions in this architecture, also referred to as extended convolutions, enlarge the receptive field of DeepLab-v3 while preserving spatial accuracy (Krestenitis et al., 2019). This innovative design enables the acquisition of baseline information from a vast geographic area. This demonstrates the model's idiosyncrasies and subtleties in difficult situations. DeepLab-v3's enhanced segmentation is facilitated by atrous convolutions (Chen et al., 2018a).

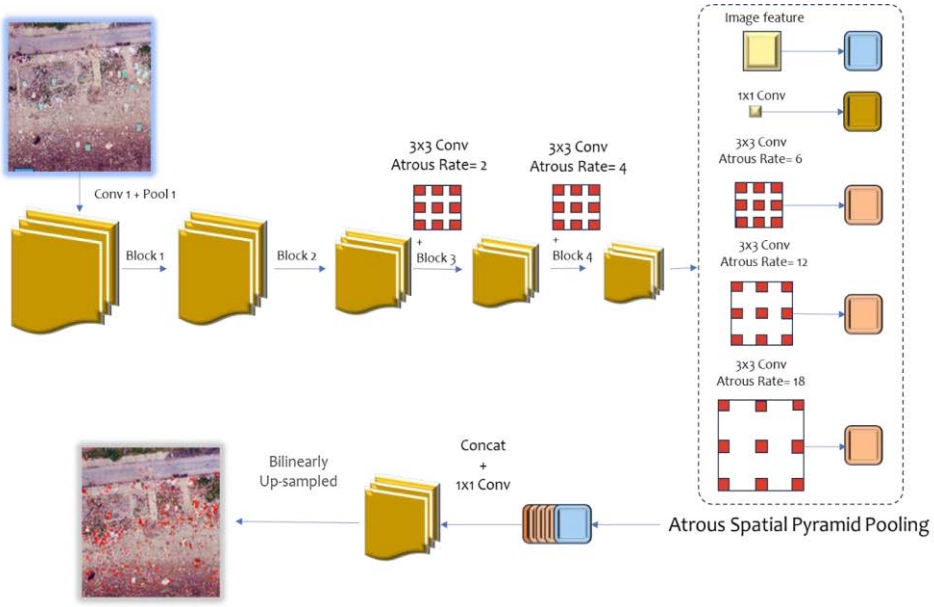

Figure 7. Deeplab-v3 architecture for debris detection using UAV data

This fundamental concept underpins the innovative Atrous Spatial Pyramid Pooling (ASPP) module in DeepLab-v3. This module effectively compiles multi-scale context (Zhang et al., 2018). As part of its mechanism, ASPP utilizes parallel atrous convolutions that operate at distinct compression levels to process environmental data at different spatial scales. This complete context aggregation approach in DeepLab-v3 improves conceptual segmentation accuracy by dividing image features of all sizes and forms.

To enhance stability and velocity, DeepLab-v3 implements numerous architectural enhancements and training methodologies (Chen et al., 2018b). The model can differentiate between coastal debris and other classes during training, thanks to class balancing. In DeepLab-v3, batch normalization and skip connections expedite closure and stabilize training dynamics. Batch normalization helps reduce the network's sensitivity to weight and bias initialization by balancing layer activation signals. Skip connections permit gradient propagation during training, which helps stabilize learning outcomes by mitigating the gradient vanishing problem. The robustness of DeepLab-v3, its fast operation, and its potential for real-world coastal-debris detection make this method very promising.

2.4.4. Design of PSPN models

The architecture of PSPN, developed by researchers at Stanford University, has been considered a modern CNN architecture (Zhu et al., 2021). Each pixel receives a specific label, and the image is organized into key

segments. The architecture shown in Figure 8 uses pyramid pooling to boost semantic segmentation while collecting multi-scale environmental information. In PSPN, the pyramid pooling module aggregates environmental features across spatial dimensions. Convolutional neural networks struggle to capture context, particularly for shaped objects and structures (Wang et al., 2022). To circumvent this issue, PSPN employs hierarchical feature aggregation and spatial pyramid pooling (SPP) to aggregate background data of varying sizes. To provide network context-aware features for coastal debris across multiple spatial levels, the pyramid pooling module divides the input feature map into a grid of sub-regions and pools data within each sub-region. PSPN may employ multi-scale feature aggregation to obtain local and global context, and subsequently separate them based on the spatial connections between the characteristics and structures of coastal debris, and compare them to other objects.

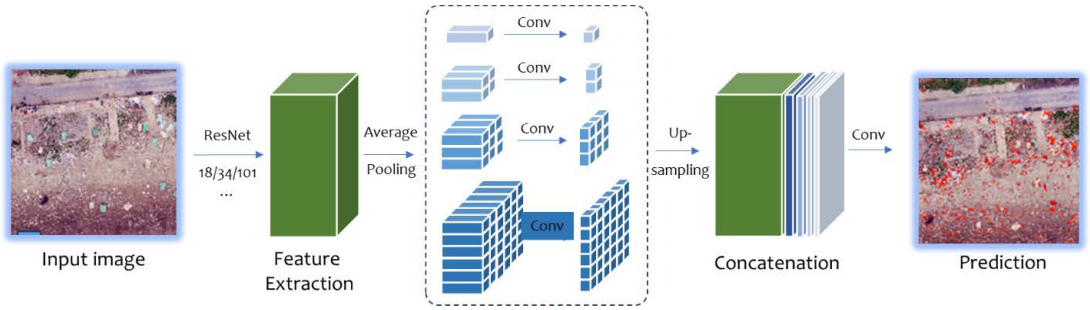

Figure 8. PSPN architecture for debris detection using UAV data

During forward propagation, the skip connections enable the network to bypass any levels (Wang et al., 2022). Meanwhile, the backpropagation gradient flows smoothly, making the PSPN one of the earliest networks to combine a pyramid-sharing module with omitted links for better segmentation performance (Zhu et al., 2021). Using PSPN, computers become capable of processing information about the environmental surroundings related to debris traces. Challenging environments become navigable, and numerous self-directed operations become feasible because PSPN builds an understanding of its surroundings through outcome models.

2.5. Techniques for model optimization

To identify debris without overfitting or underfitting, it is necessary to evaluate DL models. The projected debris outcomes were rigorously validated by comparing the total accuracy (ACC) and loss function values to the training and validation labels. All deep learning models were implemented using the TensorFlow Keras API and trained under the same experimental configuration to ensure comparability across different architectures and input sizes. The Adam optimizer was used to update model weights, with an initial learning rate of 0.001. The batch size was set to 16 for 64 × 64 and 128 × 128 input tiles, and to 8 for 256 × 256 input sub-images, to account for memory requirements. Each model was trained for up to 50 epochs. Model convergence was monitored using the validation loss, and the final model weights were selected as those yielding the lowest validation loss during training. The training process was stopped after 50 epochs when the validation loss became stable. The experiment was designed as a controlled model-comparison study rather than a full hyperparameter-optimization study. The main experimental variables were model architecture, ResNet backbone depth, and

input sub-image size, while other training parameters were kept fixed to ensure comparability among model configurations. Specifically, all models were trained using the Adam optimizer, a fixed learning rate, fixed batch sizes, and a maximum of 50 epochs. The selected models, therefore, represent the best-performing configurations within the tested experimental design, rather than globally optimized deep-learning models. No data augmentation was applied in this study; therefore, the models were trained only on the UAV-derived sub-images and their corresponding binary masks described in Section 2.3.

2.5.1. Methods for Model Evaluation

Six evaluation indicators, "accuracy," "dice," "precision," "recall," "f1," and "loss function," are employed to assess the performance of trained models. Accuracy is the ratio of the number of true predictions to the total number of data samples used in the evaluation (Dang et al., 2020b). This approach is frequently used when it is anticipated that the data classes are balanced and have equivalent sample sizes. The following formula was used to determine the accuracy value:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ samples} \quad (6)$$

Secondly, the Dice coefficient is frequently employed to compare the expected and actual results (Nguyen et al., 2025; Seale et al., 2022). The measure determines the ratio of the intersection area between the anticipated and actual outcomes to their combined areas. The model correctly segregated the data, as evidenced by a Dice score of approximately 1, with the anticipated and actual results consistent. Use the following formula to determine the value of a dice:

$$Dice = \frac{2*|M \cap N|}{|M|+|N|} \quad (7)$$

M = predicted set, N = ground truth set.

Thirdly, a model's capacity to avoid both positive and negative predictions is determined by its accuracy. The statistic estimates the ratio of accurately predicted negative instances to projected positive ones. The model's exceptional accuracy ensures that it rarely misinterprets negative situations as positive. Accuracy value determination formula:

$$Precision = \frac{Number\ of\ true\ positive\ predictions}{Number\ of\ positive\ predictions} \quad (8)$$

"Recall" is an additional critical metric for evaluating a model's ability to identify all potential positives in the dataset (Muller and Zeller, 2002; Ngo et al., 2025). The statistic is the ratio of positive instances correctly predicted to the number of positive cases in the dataset. A high recall indicates that the model is adept at identifying positive examples; however, it may commit numerous errors when predicting negative scenarios. Recall value formula:

$$Recall = \frac{Number\ of\ true\ positive\ predictions}{Total\ number\ of\ positive\ samples} \quad (9)$$

The F1 Score is a metric that assesses a DL model's ability to identify all positive predictions while minimizing false positives by combining Precision and Recall (Pouliot et al., 2019). The harmonic average of Precision and Recall is often used to assess a model that considers both variables. F1 score is calculated using the following formula:

$$F1\ score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (10)$$

Data loss due to up- and down-sampling may be induced by variations in weight, image size, and labeling (Huang et al., 2020). To mitigate input data loss, the models decreased the loss function (J). Weight optimization enhances debris predictions and minimizes information loss. The average loss values for all training sub-image datasets are as follows:

$$L = \frac{1}{m}\sum_{i=1}^{m} l^{(i)} \quad (11)$$

Here, "m" denotes the training data, while $l^{(i)}$ denotes the loss for the "i" trained sub-image. The TensorFlow Keras API was employed to generate all DL models (Gulli and Pal, 2017). The veracity of the test and

validation was verified during the training process. The DL training was terminated by coefficient convergence after 50 epochs.

To further examine model uncertainty and classification errors, a pixel-level confusion analysis was conducted by comparing the predicted debris masks with the manually annotated ground-truth mask in testing data. Both datasets were converted to binary raster format, with debris pixels assigned a value of 1 and non-debris pixels assigned a value of 0. Four confusion categories were then derived: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Based on these values, accuracy, precision, recall, F1-score, Dice coefficient, intersection over union (IoU), false positive rate (FPR), and false negative rate (FNR) were calculated. This analysis was used to identify the dominant error type and to support the interpretation of uncertainty in UAV-based coastal debris segmentation.

2.5.2. Optimizing model precision

By decreasing cost functions, optimization preserved the semantic and spatial data of coastal debris and others. SGD, Adagrad, RMSProp, Adam, and Nadam were implemented to optimize model weights and mitigate training errors (Escorcia-Gutierrez et al., 2022; Wang et al., 2017). By using continuous loss-function training-error computation, optimization is achieved. The filter weights were rapidly modified by the optimizer routines at each epoch in all DL models. The training parameters were refined through gradient-descent-guided adjustments, which identified appropriate locations to introduce new parameters. Other optimization techniques are outperformed by the Adam optimizer function in terms of accuracy and loss function value (Dang et al., 2022a, 2020b). This formula was employed to train all model parameters in the Adam optimization approaches:

$$y_{t+1} = y_t - \frac{\mu}{\sqrt{\hat{v}_t}+0.01}\hat{x}_t \quad (12)$$

In this equation, "y" denotes the parameter value, "μ" denotes the learning rate, and "t" denotes the time step (typically 10). "×" and "ϑ" are variables that estimate the first and second moments, respectively.

The performance quality of the DL model can be improved by accurately positioning the six-layer varieties described in Section 2.2.1. The authors evaluated the performance of the three main architectural networks by substituting the ResNet, DenseNet, and VGG structures. The depth and number of layers, particularly convolution layers, are the most variable. For example, ResNet34 provides UNet with 34 convolution layers (Perez et al., 2019). ResNet152 comprises 152 convolutional layers. They might be altered within trained models to optimize debris-trace extraction tasks.

2.6. Interpretation of new coastal debris

Although the DL model is trained on different data from Vietnam, the prediction task for new UAV coastal debris was performed along the coast in Giao Hai commune, Giao Thuy district, Nam Dinh province. The new input was collected from 2023 through 2025. Once the model is completed, this distinctive interpretation method necessitates no data preparation. The selection of a final model renders all requirements related to ground truth points, training data, and domain expertise unnecessary. After the new image is inserted into the model, this data is passed through the hidden-layer parameter function to convert the image data into three-dimensional band-based matrices for input. These matrices detect debris of all sizes along the coast. Numerical

classification of the input data occurs in the Fully Connected (FC) layer after the debris bounds have been delivered to it.

The authors employed drones to observe debris along the coast of Giao Hai commune. Debris information was collected both above and below ground. From above, the UAV was used to collect information within a 2 km radius of the coastline, and about 200 m from the embankment to the sea. Meanwhile, the authors arranged interviews with households living in the area affected by the flight to understand the current situation regarding debridement, collection, treatment, and concerns from local people. Information collection, supported by technology and interviews, can help the data become more detailed and multidimensional. Therefore, the authors focus on evaluating debris detection difficulties in Section 3 and on comparing our methodology with previous studies in Section 4.

3. Results

3.1. Performance of trained models

The development of DL models was deployed on 24 instances, with image sizes varying from 64, 128, and 256, and three model architectures: UNet, PSPNet, and DeepLabV3, with ResNet Conv layers varying from 18, 34, and 50. In general, the results in Table 3 show that models with Conv above 50 are unnecessary, as they consume computer resources without improving model performance. Because the performance metrics were calculated from UAV-derived sub-images, the results may be affected by spatial autocorrelation among sub-images originating from the same UAV scene. Consequently, accuracy was not used as the sole basis for model selection. The final selection also considered Dice and F1-score, validation loss, convergence stability, visual prediction quality, and the practical performance of the selected models when applied to UAV scenes not used in model training.

Table 3. Comparison of trained model performance

No.	Input size	Archi-tecture	No. Resnet Conv	Loss function of Training data	Loss function of Validating data	Accu-racy	Dice	Pre-cision	Recall	F1
1	64	UNet	18	0.13	0.16	0.93	0.51	0.7	0.52	0.6
2	64	UNet	34	0.09	0.15	0.93	0.51	0.72	0.54	0.62
3	64	UNet	50	0.12	0.17	0.93	0.58	0.71	0.49	0.58
4	64	DeepLabv3	18	0.16	0.22	0.91	0.23	0.62	0.27	0.38
5	64	DeepLabv3	34	0.15	0.22	0.91	0.19	0.68	0.21	0.31
6	64	DeepLabv3	50	0.15	0.22	0.91	0.19	0.67	0.21	0.32
7	64	PSPN	18	1.87	78.8	0.93	0.51	0.71	0.59	0.65
8	64	PSPN	34	0.09	0.15	0.93	0.54	0.73	0.51	0.59
9	64	PSPN	50	0.11	0.16	0.93	0.56	0.72	0.52	0.61
10	128	UNet	18	0.22	0.16	0.93	0.53	0.75	0.48	0.59
11	128	UNet	34	0.29	0.21	0.92	0.32	0.72	0.34	0.46
12	128	DeepLabv3	18	0.27	0.21	0.92	0.32	0.68	0.28	0.41
13	128	DeepLabv3	34	0.29	0.25	0.91	0.23	0.62	0.29	0.41
14	128	DeepLabv3	50	0.28	0.23	0.91	0.29	0.65	0.22	0.32
15	128	PSPN	18	0.28	0.17	0.93	0.47	0.76	0.46	0.57
16	128	PSPN	34	0.25	0.19	0.93	0.41	0.76	0.47	0.58
17	128	PSPN	50	0.33	0.29	0.92	0.51	0.62	0.59	0.61
18	256	UNet	18	0.17	0.19	0.92	0.36	0.75	0.35	0.48
19	256	UNet	34	0.21	0.22	0.91	0.19	0.69	0.43	0.53
20	256	DeepLabv3	18	0.24	0.26	0.59	0.29	0.59	0.25	0.35
21	256	DeepLabv3	34	0.21	0.23	0.91	0.19	0.64	0.25	0.35
22	256	DeepLabv3	50	0.23	0.24	0.91	0.32	0.62	0.31	0.41
23	256	PSPN	18	0.25	0.25	0.92	0.47	0.73	0.49	0.59
24	256	PSPN	34	0.38	0.37	0.92	0.42	0.67	0.49	0.56

The results of the model performance evaluation show that most models exhibit high convergence (Table 3). The smaller the input image size (64 × 64), the better the model identifies small debris traces, while the 256 × 256 image size can identify large debris traces. Figure 9 was used to examine the convergence behavior of the models by comparing training and validation loss curves. Only the PSPN model with a 64 × 64 image size and ResNet18 has unstable convergence, increasing the loss function index to 1.87 and 78.8 during training and validation, respectively. The PSPN-ResNet18 model with an input size of 64 × 64 showed an abnormal validation loss of 78.8, substantially higher than that of all other tested models. This value indicates an unstable or failed training run rather than a reliable model outcome. Although this model achieved relatively high accuracy and precision, its validation loss was abnormally high, and its convergence behavior was unstable. This model was not interpreted as an overfitting case based solely on accuracy or precision, but rather as an unreliable training run with poor validation behavior. Therefore, this model was excluded from the final model-selection process. The selected models were determined based on stable convergence, low validation loss, and balanced segmentation metrics.

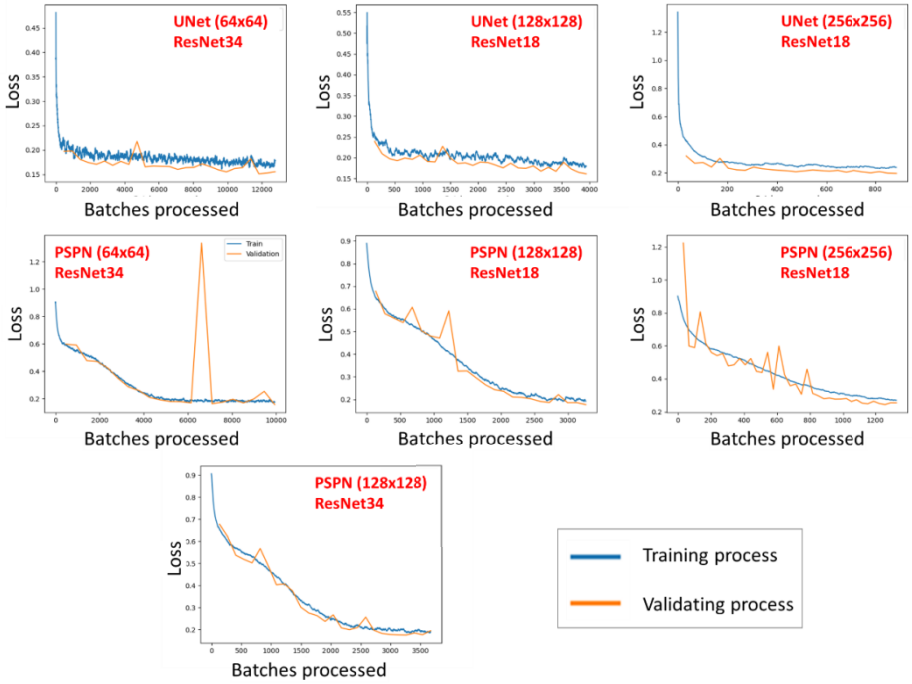

Figure 9. Training and validation loss curves of selected DL models; blue and orange lines represent training and validation loss, respectively

In addition, seven models with three different input image sizes and model architectures, with accuracy indices higher than 0.92 and precision higher than 0.72, were selected. Specifically, with an input image size of 64 × 64, the UNet-ResNet34 and PSPN-ResNet34 models are selected; with an input image size of 128 × 128, the UNet-ResNet18 and PSPN-ResNet18 and -34 models are selected; and with an input image size of 256 × 256, the UNet-ResNet18 and PSPN-ResNet18 models are selected. The

Unet-ResNet34 and PSPN-ResNet34 models with an input size of 64 × 64 were selected because they showed stable convergence, low validation loss, good segmentation metrics, and better visual detection of small and scattered coastal debris than models with larger input sizes.

Based on the loss function index convergence chart during training, the UNet models generally showed more stable convergence than the PSPN models (Fig. 10). In particular, PSPN models with a 64 × 64 image size exhibit high noise even after convergence. Models using an input image size of 256 × 256 tended to converge rapidly, in the range of 800 to 2,000 iterations. The results of the experimental interpretation of areas not included in the training set show that models using 64 × 64 input subimages were more effective at detecting small, scattered debris traces. Meanwhile, models using 128 × 128 and 256 × 256 input sizes tended to miss small debris fragments and were more sensitive to visually similar coastal objects. It reduced their dice scores compared to models using smaller input sizes.

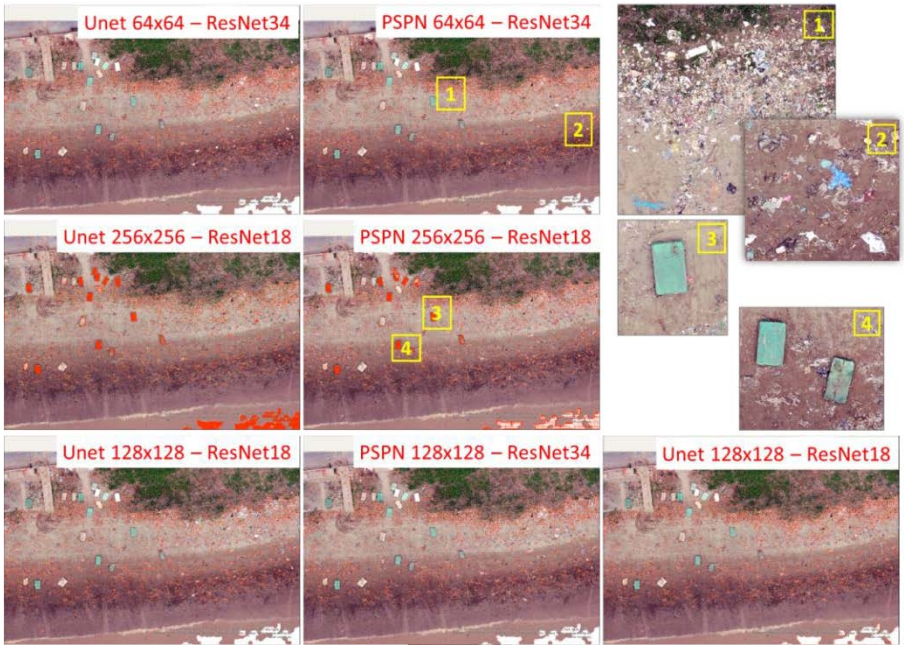

Figure 10. Visual comparison of coastal debris detection by selected DL models, red areas indicate pixels classified as coastal debris on UAV images

Foam-related materials were interpreted as a specific type of coastal debris in this study. Field observations and local interviews indicated that many foam buoys and foam fragments found on the beach were no longer in use and had been discarded by local people or during fishing activities. Therefore, these objects were classified as debris, although their origin was mainly local and land-based rather than offshore. This distinction is important for interpreting both model outputs and debris sources: the model correctly detected abandoned foam materials as coastal debris, but these detections should be understood as locally generated waste rather than marine-transported debris. Due to their bright color and texture, foam materials may still contribute to classification uncertainty, especially when mixed with sand, shells, or other light-colored coastal surfaces.

The confusion analysis (Table 4) indicates that all tested models produced relatively low

false positive rates, ranging from 0.001 to 0.002. This result shows that the models rarely classified non-debris pixels as debris under the adopted ground-truth mask. However, false-negative rates were higher, ranging from 0.218 to 0.504, indicating that omission errors were the dominant source of uncertainty. Among the tested models, UNet64-ResNet34 achieved the highest overall performance, with an accuracy of 0.975, precision of 0.986, and recall of 0.782, F1-score/Dice of 0.872, and IoU of 0.774. These results show that the main limitation of the trained segmentation models was not over-detection of debris, but the omission of some debris pixels, particularly in visually complex coastal backgrounds.

The results from Table 4 are consistent with the visual interpretation results, which show that the UNet model with a 64 × 64 input size is suitable for automatically detecting small, scattered coastal debris traces. Therefore, under the current experimental configuration, Unet-ResNet34 with 64 × 64 input size was retained as the main model for interpreting and monitoring the spatial distribution of coastal debris. This model was selected based on a combined assessment of validation loss, convergence stability, visual prediction quality, confusion-matrix-based metrics, and the ability to detect small, scattered debris. Therefore, the distribution of coastal debris was interpreted and monitored over four periods in the Giao Hai coastal area, Nam Dinh province, in the next section.

Table 4. Confusion-based performance metrics of DL segmentation models for testing data

Model	Accuracy	Precision	Recall	F1-score	Dice	IoU	FPR	FNR
UNet64-ResNet34	0.975	0.986	0.782	0.872	0.872	0.774	0.001	0.218
UNet128-ResNet18	0.949	0.977	0.549	0.703	0.703	0.542	0.002	0.451
UNet256-ResNet18	0.946	0.969	0.524	0.68	0.68	0.515	0.002	0.476
PSPN128-ResNet18	0.95	0.974	0.557	0.709	0.709	0.549	0.002	0.443
PSPN128-ResNet34	0.955	0.976	0.6	0.743	0.743	0.591	0.002	0.4
PSPN64-ResNet34	0.952	0.974	0.574	0.722	0.722	0.566	0.002	0.426
PSPN256-ResNet18	0.943	0.967	0.496	0.655	0.655	0.487	0.002	0.504

3.2. Distribution of debris along the coast in Giao Thuy district, Nam Dinh province, Vietnam

The research team collected UAV data at four times: March 2023 *(dry season, before and after debris collection with the support of people and volunteer teams),* September 2023 *(rain season),* and March 2024 *(end of the dry season).* In March 2023, images were collected in March 2023 with a range of about 200 m along the coast, providing effective information about the debris collection work of local volunteer teams in preserving and protecting the regional landscape and environment. The next images, taken in September 2023 and March 2024, are to survey and compare the return of debris after 1 year. After resampling the data to a resolution of 0.008 m, the data is fed into a DL model to estimate the distribution of coastal debris.

The results show that areas with a high concentration of debris are located in the area with the road leading to the dyke from the Commune People's Committee, to the sea, and gradually decreasing to both sides along the coast. Nylon and bags are buried in the soil layer, making it difficult to identify accurately. With UAV image data, the study focused on identifying traces of debris exposed on the surface and calculating the coverage of this type of debris along the coast (Figs. 11 and 12). Debris at the survey site in March (dry season) is less than in September (rainy season). The volunteer team's debris collection work, despite leaving a small

number of bags along the coast, has significantly reduced the area of debris compared to before the collection. The study assessed the coverage of coastal debris in a 200-square-meter test area. The results showed that the debris cover before debris collection in March 2023 was nearly 15%. After collection, the debris cover reduces significantly to nearly 0.9%. However, when the rainy season came, it increased back to 12% in September 2023 and only just over 6% in March 2024.

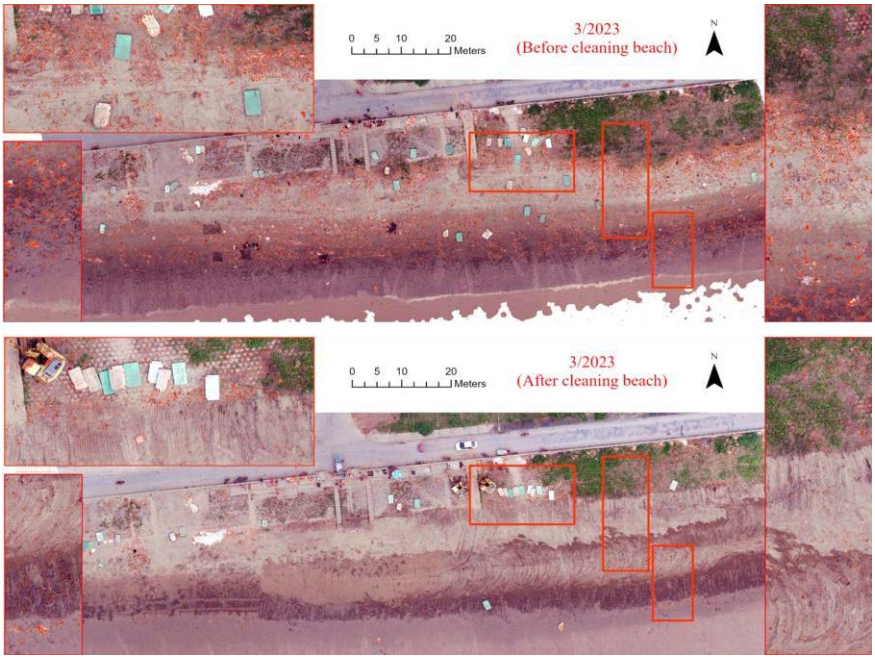

Figure 11. Coastal debris distribution before and after cleaning activities in March 2023; red areas indicate coastal debris detected using the PSPN-Resnet34 model with a 64×64-pixel input size

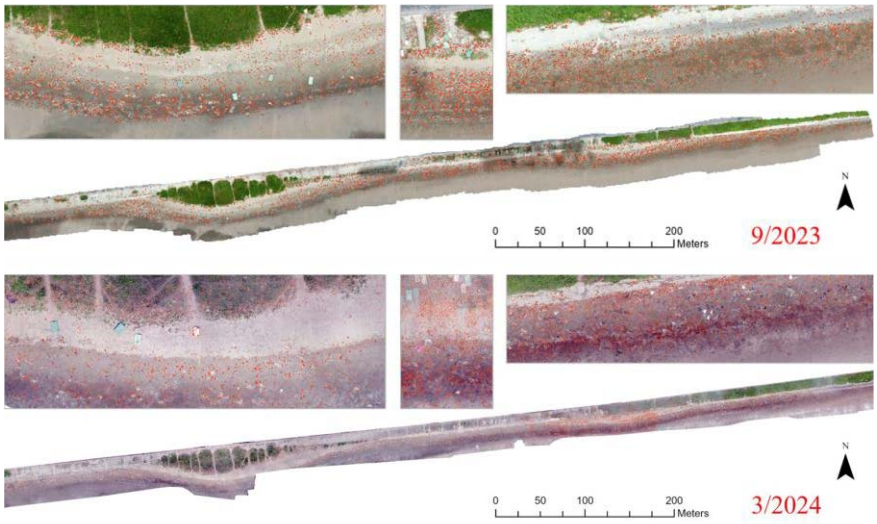

Figure 12. Seasonal variation in coastal debris distribution along the Giao Hai coast; red areas indicate coastal debris detected using the PSPN-Resnet34 model with a 64 × 64-pixel input size

The interpretation results for the four periods show that the DL model identified most of the areas where debris appears along the coast of the study area, especially scattered debris. However, in areas with high debris concentration, such as the bottom of the embankment, the phenomenon of transmission noise interpretation often occurs. Here, the debris is concentrated, piled up, and chaotic, making the interpretation results not cover the entire area of debris. According to fieldwork, debris is concentrated in many types, making it difficult for the model to distinguish between debris and other types. It can be clearly observed in the interpretation results of images collected in September 2023, the rainy season, when debris is brought in from the sea in large quantities.

As a result, the selected DL models have completely separated people's foam buoys from other coastal debris along the coast. Bags were mainly found along the coast of Giao Hai commune, whereas bottles are almost nonexistent or account for a very small percentage, unlike on islands in Vietnam. It is analyzed more clearly in the following sections.

3.3. Distribution of debris from the inland to the offshore area

The quantity of coastal debris in Giao Hai commune was estimated from images acquired on four separate occasions. It is evident that a significant amount of debris was concentrated in the tidal area and the foot of the embankment in March 2023, before its collection. The debris was distributed over a distance of 2 km along the horizontal littoral, approximately 200 m from the foot of the embankment to the sea. After the collection process in March 2023 - a month during the dry season - the quantity of coastal debris has progressively decreased. Nevertheless, the quantity of debris in this location resurfaced in September 2023 and March 2024 - months during the rainy season.

A combination of surveys and experiments was conducted with 30 families and local officials in the coastal areas of Giao Hai commune and the Con Lu coastline in Xuan Thuy National Park. The analyzed data indicated problems within the public information system regarding government management, as well as the reception and reflection processes. The outcomes showed that about 73% of respondents observed coastal debris above the embankment, and another 89% observed it on sandy coasts (Fig. 13). Additionally, 11% of the debris was found on the sea floor, and 52% was floating on the surface. The residents of Giao Hai commune state that coastal debris has existed for a long time, yet has shown a sharp increase over the last ten years. The people in the study area identified nylon bags, bottles and food containers, polystyrene crates, and fishing nets (Fig. 14). Additional debris found in the area includes small foam products, items, and wood materials, including lumber structures.

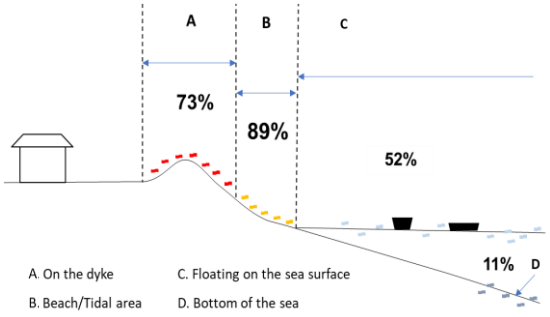

Figure 13. Distribution of debris from the inland to offshore areas based on the DL model, UAV data, and interview process

Interviews confirmed the operation of two debris collection units that actively remove debris from coastal areas in Giao Hai commune, Giao Thuy district, Nam Dinh province. Environmentalist groups, together with community volunteers, student organizations, and individual members, collaborate to conduct debris collection activities during major environmental

awareness days, including "Earth Overload Day," "Sea Week," and "Islands". The Commune People's Committee, together with cooperatives, organized gathering teams that have served as small independent teams. There are only two or three yearly collection drives that fall short of their communication objectives.

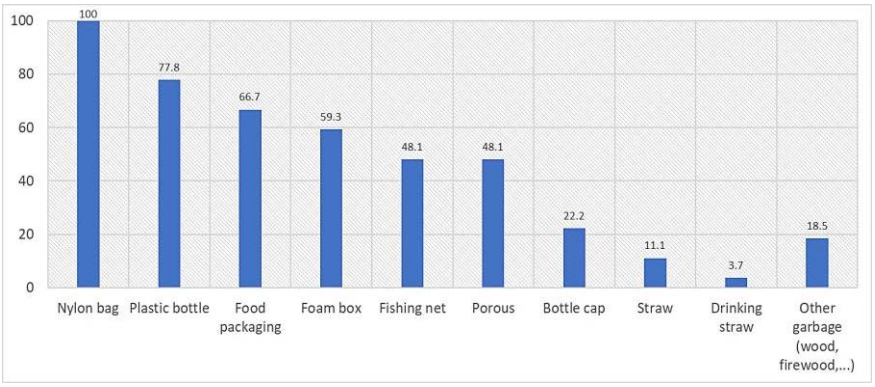

Figure 14. Debris types found along the coast of Giao Hai district, Nam Dinh province

Collected coastal debris is directed to a main facility for treatment. The debris is incinerated, then buried at the designated local temporary landfill. The commune collects the debris and deposits it at the temporary debris collection sites. Failing to use either technological or manual debris-processing methods, including improper burial and un-technological burning practices, poses environmental and health risks. The authors discovered that environmental cleanliness promotion and defense activities had been sporadic at the site. Regular promotion and public education on coastal debris collection, its transportation, and subsequent treatment must be conducted by local authorities. Appropriate awareness measures can make environmental protection education and coastal trash management practices more accessible.

4. Discussions

4.1. Causes and origins of coastal debris

This study emphasized the need for a spatial approach to the shoreline, rather than a discrete single-site survey of debris. Once managers know the accumulation of coastal debris, they can apply the developed. Unet-ResNet34 with a 64 × 64 input size, trained on UAV data, to detect debris along the coast. Contrary to the view that debris is mainly carried away by wind or ocean currents, in Nam Dinh, according to Do et al. (2025),, the majority of debris accumulates within the 0–10 m depth range. Coastal debris distribution may reflect the interaction between shoreline characteristics, waves, tides, and longshore currents. Areas with complex shorelines, interspersed with wave breakers such as intertidal areas, sand bars, or artificial embankments, are often hotspots for debris accumulation, forming localized pollution zones. A strong driver is the longshore current - a parallel water flow formed by waves hitting the shore obliquely. This current can transport debris along the shoreline for hundreds of meters to kilometers, especially in estuarial areas such as from the Ba Lat estuary southwards. When encountering blocking factors such as sandy coasts, rocky beaches, or wave-breaking structures, debris moving as special sediments loses kinetic energy, settles, and accumulates there. This phenomenon is clearly recorded along the Giao Hai and Giao Phong coasts, where effective current-blocking formations are present.

Based on the above analysis, depending on the characteristics of the coastal type, the ability to accumulate debris can be different as follows:

- Sandy accumulated coasts: with a slight slope and a concave shape, allowing waves to push debris into the tidal zone and then get stuck. In the rainy season, waves from the Gulf of Tonkin and the Northeast wind bring debris from the estuarial area to nearby shores. Heavier debris can settle at the base or edges of the sand dunes, while lighter debris can be pushed higher up, sometimes behind the dunes.

- Sand bars: Formed by the process of sedimentation in places where the current is weak, often with calm water inside lagoons. It is an ideal place for debris to drift in and accumulate. Debris is often stuck at the base of the sand bar - where the current and waves converge, reducing energy.

- Estuarial areas: Where River and sea currents meet, the flow rate decreases and sediments are deposited strongly. Debris discharged from the river often gets stuck in alluvial flats, shallow water areas, or in mangrove roots. It is an area where long-lasting natural "debris traps" form easily.

These coast types act as a natural "biological-physical trap", where debris is accumulated or dispersed depending on the morphology and impact dynamics. Three coastal types are formed above the sea level. It enables the application of UAVs and the development of DL models that can easily detect debris on the ground. Identifying and zoning types of coastlines likely to accumulate debris is an important basis for building risk maps, supporting the monitoring and effective collection of debris in coastal areas sensitive to tourism, fisheries, and biodiversity.

The geomorphological interpretation in this study should be understood as a field- and UAV-based spatial interpretation rather than a quantitative geomorphic-control analysis. Coastal features such as embankments, tidal flats, sand bars, estuarine areas, and longshore-current pathways were used to support the interpretation of observed debris accumulation patterns. Future studies should integrate UAV-derived DEMs, shoreline morphology classification, wave and tidal data, current patterns, elevation and slope metrics, and distance-based variables to quantify how coastal landforms and hydrodynamic processes control debris transport, trapping, and accumulation.

4.2. Proposed 4.0 technology framework for coastal debris management

The empirical contribution of this study is the UAV/deep-learning-based detection and mapping of coastal debris. Based on these results and adapted from studies of Nunkhaw and Miyamoto (2024) and Schmidt et al. (2023), a 4.0 technology-based framework is proposed as a management recommendation, not as a fully implemented or field-tested operational system. The framework links four main components: (1) community reporting, (2) UAV-based verification, (3) AI-assisted debris mapping, and (4) cleanup planning and post-cleaning assessment. In this framework, residents report debris hotspots through a smartphone application or digital platform. Local authorities screen the reports and coordinate UAV surveys where necessary. The trained DL model then processes UAV images to verify debris locations, map debris distribution, and provide spatial evidence for management decisions. After cleanup activities, UAV monitoring can be repeated to assess removal effectiveness and update debris maps.

A limitation of this study is that the proposed 4.0 technology-based management framework has not yet been implemented as an operational system in the field. The empirical contribution of this paper is the

development and evaluation of UAV-based deep-learning models for coastal debris detection and mapping. The management framework is therefore presented as a recommended application pathway that requires further testing, institutional coordination, and community participation before operational deployment. The framework requires basic digital infrastructure, UAV equipment, trained operators, AI-based image-processing capacity, institutional coordination, and a modest budget for repeated monitoring. Compared with conventional field surveys, UAV-AI monitoring may reduce field time and improve spatial coverage, but its operational effectiveness still requires field testing.

This framework should therefore be regarded as an application pathway derived from the modeling results. Future studies should not only test the operational feasibility of the 4.0 framework, but also strengthen the geomorphological interpretation of coastal debris. Further research is needed to clarify the sources of debris, transport pathways, shoreline trapping mechanisms, and the roles of coastal landforms, waves, tides, longshore currents, embankments, tidal flats, and sand bars in controlling debris accumulation and redistribution.

5. Conclusions

This research has shown great potential for combining UAV technology and DL algorithms to supervise, identify, and control coastal debris. The system detected complex debris along the coastline of Vietnam with high accuracy by deploying UNet-ResNet34 and PSPNet-ResNet34 on an input image size of a 64-pixel grid, which not only applied to any region of interest but also enabled evaluation of whether debris density varied across time and space. Field findings indicate that debris tends to accumulate most on shore types with barrier terrain, such as embankments, tidal areas, sand bars, and estuarial areas, which are affected by longshore currents, waves, and seasonal conditions. The survey results also show that debris removal effectiveness is limited without coordination between the community and local authorities.

This study demonstrates the potential of UAV imagery and deep-learning models for detecting and mapping coastal debris. The selected models provided useful spatial information on debris distribution and temporal change along the Giao Hai coast. Based on these empirical findings, a 4.0 technology-based framework is proposed as a future management application that links community reporting, UAV monitoring, AI-based mapping, and cleanup evaluation. However, this framework remains a recommendation and requires further field implementation and institutional testing before it can be considered an operational management system.

Acknowledgements

The paper was conducted under the national project funded by the Ministry of Science and Technology of Vietnam, code KC.09.02/21-30.

References

Ayub Y., Ren J., 2024. Co-Pyrolysis of biomass and plastic waste: Process prediction and optimization based on Artificial Intelligence and response optimizer surrogate model. Process Saf. Environ. Prot., 186, 612–624. https://doi.org/10.1016/j.psep.2024.04.049.

Berto M.R.S., Reis M., Lima T.A.S., Soares C.N., Silva I.M. dos S., Cavalcanti A.S., Pavão J.M., 2025. Mapping the seasonality of marine debris waste in a tropical touristic hotspot. J. Coast. Conserv., 29, 25. https://doi.org/10.1007/s11852-025-01109-x.

Chen L.C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L., 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous

Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell., 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

Chen L.C., Zhu Y., Papandreou G., Schroff F., Adam H., 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. Computer Vision and Pattern Recognition, arXiv:1802.02611v3, 18. https://doi.org/10.48550/arXiv.1802.02611.

Choi J., Lim B., Yoo Y., 2023. Advancing Plastic Waste Classification and Recycling Efficiency: Integrating Image Sensors and Deep Learning Algorithms. Appl. Sci., 13(18), 10224. https://doi.org/10.3390/app131810224.

Dang K.B., Dang V.B., Bui Q.T., Nguyen V.V., Pham T.P.N., Ngo V.L., 2020a. A Convolutional Neural Network for Coastal Classification Based on ALOS and NOAA Satellite Data. IEEE Access, 8, 11824–11839. https://doi.org/10.1109/ACCESS.2020.2965231.

Dang K.B., Dang V.B., Ngo V.L., Vu K.C., Nguyen H., Nguyen D.A., Nguyen T.D.L., Pham T.P.N., Giang T.L., Nguyen H.D., Hieu Do T., 2022a. Application of deep learning models to detect coastlines and shorelines. J. Environ. Manage, 320, 115732. https://doi.org/10.1016/j.jenvman.2022.115732.

Dang K.B., Ha T., Nguyen T., Nguyen H.D., Truong Q.H., Vu T.P., Pham H.N., Duong T.T., Giang V.T., Nguyen D.M., Bui T.H., Burkhard B., 2022b. U-shaped deep-learning models for island ecosystem type classification, a case study in Con Dao Island of Vietnam. One Ecosyst., 7, 23. https://doi.org/10.3897/oneeco.7.e79160.

Dang K.B., Ngo C.C., Ngo V.L., Dang V.B., 2022c. Coastal Vulnerability Assessment in Son Tra - Cua Dai region. VNU J. Sci. Earth Environ. Sci., 38(3), 55–65. https://doi.org/10.25073/2588-1094/vnuees.4842.

Dang K.B., Nguyen M.H., Nguyen D.A., Phan T.T.H., Giang T.L., Pham H.H., Nguyen T.N., Van Tran T.T., Bui D.T., 2020b. Coastal wetland classification with deep U-Net convolutional networks and Sentinel-2 imagery: A case study at the Tien Yen estuary of Vietnam. Remote Sensing, 12(19), 3270. https://doi.org/10.3390/rs12193270.

Diakogiannis F.I., Waldner F., Caccetta P., Wu C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. Remote Sens., 162, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013.

Do T.N., Dang K.B., Nguyen T.Y., Pham T.P.N., 2025. Analyzing the Relationship between Coastal Geomorphological Features and Coastal Litter Accumulation in Nam Dinh Area. VNU J. Sci. Earth Environ. Sci., 41(4). https://doi.org/10.25073/2588-1094/vnuees.5333.

Escorcia-Gutierrez J., Gamarra M., Beleño K., Soto C., Mansour R.F., 2022. Intelligent deep learning-enabled autonomous small ship detection and classification model. Comput. Electr. Eng., 100, 107871. https://doi.org/10.1016/j.compeleceng.2022.107871.

Fu Y., Deng J., Wang H., Comber A., Yang W., Wu W., You S., Lin Y., Wang K., 2021. A new satellite-derived dataset for marine aquaculture areas in China's coastal region. Earth Syst. Sci. Data, 13(5), 1829–1842. https://doi.org/10.5194/essd-13-1829-2021.

Giuliani S., Bellucci L.G., Nhon D.H., 2019. The Coast of Vietnam: Present Status and Future Challenges for Sustainable Development, Second Edi. ed, World Seas: An Environmental Evaluation. Elsevier Ltd., 415–435. https://doi.org/10.1016/B978-0-08-100853-9.00027-0.

Gulli A., Pal S., 2017. Deep Learning with Keras - Implement neural networks with Keras on Theano and TensorFlow. Birmingham, UK : Packt Publishing, ISBN: 9781787128422.

Huang H., Lin L., Tong R., Hu H., Zhang Q., Iwamoto Y., Han X., Chen Y.W., Wu J., 2020. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2020-May, 1055–1059. https://doi.org/10.1109/ICASSP40776.2020.9053405.

Iqbal U., Barthelemy J., Perez P., Davies T., 2022. Edge-Computing Video Analytics Solution for Automated Plastic-Bag Contamination Detection: A Case from Remondis. Sensors (MDPI), 22(20), 7821. https://doi.org/10.3390/s22207821.

Jafarzadeh H., Mahdianpari M., Gill E., Mohammadimanesh F., Homayouni S., 2021. Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. Remote Sensing, 13(21), 4405. https://doi.org/10.3390/rs13214405.

Krestenitis M., Orfanidis G., Ioannidis K., Avgerinakis K., Vrochidis S., Kompatsiaris I., 2019. Oil spill identification from satellite images using deep

neural networks. Remote Sens., 11(15), 1762. https://doi.org/10.3390/rs11151762.

Kubiszewski I., Anderson S.J., Costanza R., Sutton P.C., 2016. The Future of Ecosystem Services in Asia and the Pacific. Asia Pacific Policy Stud., 3(3), 389–404. https://doi.org/10.1002/app5.147.

Luo W., Han W., Fu P., Wang H., Zhao Y., Liu K., Liu Y., Zhao Z., Zhu M., Xu R., Wei G., 2022. A Water Surface Contaminants Monitoring Method Based on Airborne Depth Reasoning. Processes, 10(1), 131. https://doi.org/10.3390/pr10010131.

Matching F., Yasir M., Liu S., Xu M., Aguilar F.J., Wan J., Wei S., Pirasteh S., Member S., Fan H., Islam Q.U., 2026. TFST : Two-Frame Ship Tracking for SAR Using. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 19, 3175–3189. https://doi.org/10.1109/JSTARS.2025.3647680.

Mathis J., Finder S., Werner M., Maddalene T., Nguyen T.Y.T., Nguyen V.C., Jambeck J., 2021. Circularity Assessment Protocol - Nam Dinh, Vietnam. University of Georgia, Athens, GA, USA.

Merlino S., Paterni M., Locritani M., Andriolo U., Gonçalves G., Massetti L., 2021. Citizen science for marine litter detection and classification on unmanned aerial vehicle images. Water (Switzerland), 13(23), 3349. https://doi.org/10.3390/w13233349.

Molnár A.D., Málnás K., Bőhm S., Gyalai-Korpos M., Cserép M., Kiss T., 2024. Comparative Analysis of Riverine Plastic Pollution Combining Citizen Science, Remote Sensing and Water Quality Monitoring Techniques. Sustainability, 16(12), 5040. https://doi.org/10.3390/su16125040.

Muller D., Zeller M., 2002. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. Agric. Econ., 27(3), 333–354. https://doi.org/10.1111/j.1574-0862.2002.tb00124.x.

Nam P.N., Pham Q.T., Ngo Thi X.T., Nguyen Thi M.D., Duong T.T., Le Thi P.Q., Le N. Da, Phuong N.A., Duong T.N., Le Q.T., 2022. Initial survey on microplastic waste in coastal water in Nam Dinh. Vietnam J. Mar. Sci. Technol., 22(2), 209–216. https://doi.org/10.15625/1859-3097/16544.

Ngo V.L., Nguyen H., Dang K.B., Giang T.L., Dang V.B., Do T.H., Nguyen M.H., Dang N.V., Dao M.D., 2025. Advancing debris flow detection based on deep learning model and high-resolution images. Vietnam J. Earth Sci., 47(2), 290–314. https://doi.org/10.15625/2615-9783/23027.

Nguyen H., Dang K.B., Giang T.L., Dang V.B., Dang N.V., Dao M.D., 2025. Exploring Various Deep Learning Models for High-Precision Landslide Tracing in Very-High Resolution Remote Sensing Imagery. J. Indian Soc. Remote Sens., 1–23. https://doi.org/10.1007/s12524-025-02267-z.

Nguyen T.T., Ngo H.H., Guo W., Nguyen H.Q., Luu C., Dang K.B., Liu Y., Zhang X., 2020. New approach of water quantity vulnerability assessment using satellite images and GIS-based model: An application to a case study in Vietnam. Sci. Total Environ., 737, 139784. https://doi.org/10.1016/j.scitotenv.2020.139784.

Nunkhaw M., Miyamoto H., 2024. An Image Analysis of River-Floating Waste Materials by Using Deep Learning Techniques. Water (Switzerland), 16(10), 1373. https://doi.org/10.3390/w16101373.

Perez H., Tah J.H.M., Mosavi A., 2019. Deep learning for detecting building defects using convolutional neural networks. Sensors (Switzerland), 19(16), 3556. https://doi.org/10.3390/s19163556.

Pham T.T., Dang K.B., Giang T.L., Hoang T.H.N., Le V.H., Ha H.N., 2024. Deep learning models for monitoring landscape changes in a UNESCO Global Geopark. J. Environ. Manage, 354, 120497. https://doi.org/10.1016/j.jenvman.2024.120497.

Pouliot D., Latifovic R., Pasher J., Duffe J., 2019. Assessment of convolution neural networks for wetland mapping with landsat in the central Canadian boreal forest region. Remote Sens., 11(7), 772. https://doi.org/10.3390/rs11070772.

Pucino N., Kennedy D.M., Young M., Ierodiaconou D., 2022. Assessing the accuracy of Sentinel-2 instantaneous subpixel shorelines using synchronous UAV ground truth surveys. Remote Sens. Environ., 282, 113293. https://doi.org/10.1016/j.rse.2022.113293.

Sakti A.D., Rinasti A.N., Agustina E., Diastomo H., Muhammad F., Anna Z., Wikantika K., 2021. Multi-scenario model of plastic waste accumulation potential in indonesia using integrated remote sensing, statistic and socio-demographic data. ISPRS Int. J. Geo-Information, 10(7), 481. https://doi.org/10.3390/ijgi10070481.

Schmidt T., Kuester T., Smith T., Bochow M., 2023. Potential of Optical Spaceborne Sensors for the Differentiation of Plastics in the Environment. Remote Sens., 15(8), 2020. https://doi.org/10.3390/rs15082020.

Seale C., Redfern T., Chatfield P., Luo C., Dempsey K., 2022. Coastline detection in satellite imagery: A deep learning approach on new benchmark data. Remote Sens. Environ., 278, 113044. https://doi.org/10.1016/j.rse.2022.113044.

Sivapriya M.S., Suresh S., 2023. ViT-DexiNet: a vision transformer-based edge detection operator for small object detection in SAR images. Int. J. Remote Sens., 44(22), 7057–7084. https://doi.org/10.1080/01431161.2023.2277167.

Stoian A., Poulain V., Inglada J., Poughon V., Derksen D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. Remote Sens., 11(17), 1986. https://doi.org/10.3390/rs11171986.

Suhail B., Kanwal H., Arsalan M., 2023. Composition and relative abundance of plastic debris along Manora and Sandspit beaches, Karachi, Pakistan. J. Coast. Conserv., 28, 9. https://doi.org/10.1007/s11852-023-01006-1.

Thi Nham T.N., 2024. Plastic Waste and Difficulties in Management in Vietnam. Eng. Technol. J., 9, 3348–3352. https://doi.org/10.47191/etj/v9i01.14.

Tran T.T.T., Pham H.K., Nguyen H.M., 2020. Assessing the current status of rural domestic solid waste management in Nam Dinh province. J. Min. Earth Sci., 61(6), 82–89. https://doi.org/10.46326/jmes.2020.61(6).09.

Tsai F.M., Bui T.D., Tseng M.L., Lim M.K., Tan R.R., 2020. Sustainable solid-waste management in coastal and marine tourism cities in Vietnam: A hierarchical-level approach. Resour. Conserv. Recycl., 1168, 05266. https://doi.org/10.1016/j.resconrec.2020.105266.

Tsai T.H., Tseng Y.W., 2023. BiSeNet V3: Bilateral segmentation network with coordinate attention for real-time semantic segmentation. Neurocomputing, 532, 33–42. https://doi.org/10.1016/j.neucom.2023.02.025.

Wahid N.M., Mutaqin B.W., 2024. Tidal fluctuation effect on marine debris characteristics in the Kulon Progo beaches of Yogyakarta, Indonesia. J. Coast. Conserv., 28, 37. https://doi.org/10.1007/s11852-024-01036-3.

Wang L., Yang Y., Min R., Chakradhar S., 2017. Accelerating deep neural network training with inconsistent stochastic gradient descent. Neural Networks, 93, 219–229. https://doi.org/10.1016/j.neunet.2017.06.003.

Wang Xi, Guo Y., Wang S., Cheng G., Wang Xinquan, He L., 2022. Rapid detection of incomplete coal and gangue based on improved PSPNet. Meas. J. Int. Meas. Confed., 201, 111646. https://doi.org/10.1016/j.measurement.2022.111646.

WWF, 2023. Report on plastic waste generation in 2022. World Wide Fund for Nature. Thanh Nien Publishing House.

Yasir M., Liu S., Pirasteh S., Xu M., Sheng H., Wan J., de Figueiredo F.A.P., Aguilar F.J., Li J., 2024. YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8. Int. J. Appl. Earth Obs. Geoinf., 134, 104137. https://doi.org/10.1016/j.jag.2024.104137.

Yasir M., Liu S., Xu M., Sheng H., Aguilar F.J., do Lago Rocha R., de Figueiredo F.A.P., Isiacik Colak A.T., Hossain M.S., 2026. SSGNet: A single-source generalization model for ship detection in UAV imagery under challenging maritime environments. Ocean Eng., 348, 124120. https://doi.org/https://doi.org/10.1016/j.oceaneng.2025.124120.

Zhang P., Ke Y., Zhang Z., Wang M., Li P., Zhang S., 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. Sensors (Switzerland), 18(11), 3717. https://doi.org/10.3390/s18113717.

Zhu X., Cheng Z., Wang S., Chen X., Lu G., 2021. Coronary angiography image segmentation based on PSPNet. Comput. Methods Programs Biomed., 200, 105897. https://doi.org/10.1016/j.cmpb.2020.105897.

Zia H., Jawaid M.U., Fatima H.S., Hassan I.U., Hussain A., Shahzad S., Khurram M., 2022. Plastic Waste Management through the Development of a Low Cost and Light Weight Deep Learning Based Reverse Vending Machine. Recycling, 7(5), 70. https://doi.org/10.3390/recycling7050070.