scientific reports

Check for updates

OPEN

Species habitat modeling based on image semantic segmentation

Lingjun Wang[1,2], Haofeng Tan[3], Peng Luo[4,5], Liqiu Meng[4] & Teng Fei[1,6]✉

Habitat monitoring has emerged as a crucial practice for preserving ecological environments and ensuring species reproduction. Traditional habitat modeling often relies on the "lasagna model"—a McHarg-style approach that focuses on the ecological niche formed by the combined effect of multiple geographical factors at a single location. This model, however, overlooks the influence of the broader surrounding environment on habitat suitability. In this study, we propose a habitat modeling framework that integrates surrounding environmental conditions by employing kernel density analysis and a semantic segmentation method. The results demonstrate that kernel density analysis is effective in expanding the presence-only data into presence-absence data for habitat modeling. The semantic segmentation method, Segformer, outperforms the traditional MaxEnt in mapping the habitat of the Sandpiper family in Taiwan, achieving a higher Area Under the Curve (AUC) score (0.76 vs. 0.69). Another case study of the Swallow family indicates the limitations of the proposed method. This study highlights the potential of applying deep learning methods to habitat modeling, contributing to more comprehensive biodiversity assessments and conservation planning.

Keywords Species distribution models, Kernel density analysis, Deep learning, Semantic segmentation

Habitat loss and fragmentation have been major drivers and the greatest threats to biodiversity in ecosystems over the last century[1]. Studies have reported a biodiversity loss ranging from 13 to 75% across five continents[2]. In response, since 1990, the European Union has launched more than 150 habitat-related projects to promote habitat management and restoration, each with a budget of over 80 million Euros[3]. The substantial investment of human and financial resources underscores the importance of research on species habitats.

To support such conservation efforts, species distribution models (SDM) have been widely used to identify and assess potential habitats[4]. It helps to explore the spatial distribution of species by transforming environmental variables into either probabilities of species presence or binary presence-absence outcomes at a specific location.

Previous methods

SDMs typically rely on two key data sources: species observation records and environmental variables. Climate, topography, soil type, land-use type, and distance to specific land-use type are recognized as important environmental variables[5]. Building on these variables, previous research has developed a wide range of SDMs, generally classified into: traditional statistical methods, machine learning-based methods, and presence-only methods.

Traditional statistical methods often treat species distribution modeling as a classification or regression problem, using environmental variables as input and predicting either the probability of species presence or a binary outcome indicating presence or absence at a given location. Early methods, such as the generalized linear model (GLM)[6] and logistic regression[7], assume a linear relationship between environmental variables and species distribution, limiting their ability to model non-linear relationships. Generalized additive model (GAM)[8] and multiple adaptive regression splines (MARS)[9] have been introduced to solve this issue. Additionally, as spatial autocorrelation in SDMs gained increasing attention, spatial analysis techniques have been integrated into these methods to enhance model performance[10].

Despite these advances, traditional statistical methods struggle to model more complex relationships. Machine learning techniques, particularly supervised classification methods, have been increasingly adopted in SDMs, including decision trees[11], support vector machines (SVMs)[12], and boosted regression trees (BRTs)[13].

[1]School of Resource and Environmental Sciences, Wuhan University, No.129, Luoyu Road, Wuhan 430079, Hubei, China. [2]School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany. [3]Department of Geography, University of South Carolina, Columbia 29208, USA. [4]Chair of Cartography and Visual Analytics, Technical University of Munich, 80333 Munich, Germany. [5]Environmental Change Institute, University of Oxford, Oxford OX1 3QY, UK. [6]School of Forestry, University of Canterbury, Christchurch 8140, New Zealand. ✉email: feiteng@whu.edu.cn

Decision trees are especially effective for capturing hierarchical relationships between environmental variables, often producing more ecologically interpretable results than other machine learning methods[14].

Traditional statistical and machine learning-based SDMs require both positive (presence) and negative (absence) species observation records. However, negative records are often difficult to collect, as it is challenging to confirm the absence of species at a specific location. To address this limitation, the third type of SDMs has been developed specifically for presence-only records. These methods are two folds: one directly uses presence-only data, such as maximum entropy (MaxEnt), which demonstrates powerful performance even when based on limited presence records[15]. The second approach generates pseudo-absence records from modeling based on prior expert knowledge[16,17], which introduces subjective bias and may significantly influence the accuracy of species distribution predictions. Additionally, the imbalance in the amount of presence and pseudo-absence records can lead to biased modeling[18,19].

Deep learning involved in SDMs

Recent advances in deep learning offer new opportunities for SDMs. Deep learning has been applied in ecological studies since 2000, with increasing adoption after 2016, alongside the development of deep learning networks such as convolutional neural networks (CNNs), recurrent neural networks (RNNs)[20].

One key advantage of deep learning is its capability to model complex and non-linear relationships between species and environmental variables through multiple hidden neural network layers. In addition, deep learning methods for image-based tasks can effectively capture broader spatial information, which is often ignored in traditional SDMs. For example, in an image scene classification task, a deep learning model takes an image as input to predict a scene category, such as sidewalk or forest. If only a small portion of the image is taken as input (e.g., trees), the model may misclassify the overall scene. This idea can be extended to habitat modeling: predictions that only focus on environmental variables at a single location, while ignoring broader surrounding conditions, may lead to inaccurate results. To address this limitation, previous studies have applied image classification-based methods to model species distribution[3,21,22]. Recently, semantic segmentation methods have demonstrated strong performance in pixel-level classification. Each pixel within the image is classified into different categories, like trees or buildings, by semantic segmentation methods, considering the characteristics of surrounding pixels collectively. Despite their potential, the application of this kind of deep learning methods remain unexplored in SDMs.

Proposed method

There are two major research gaps observed in existing SDM studies. First, most available species observation data sources only provide presence records, whereas many SDMs, including traditional statistical and machine learning-based methods, require presence-absence data. While prior expert knowledge can be used to generate pseudo-absence records, this approach lacks a reliable method to create balanced and unbiased absence datasets. Second, the environmental variables exert influence over a range of areas. The ecological niche of an area is not only shaped by the local environment, but also by the surrounding environment. However, traditional SDMs only consider site-specific environmental variables, overlooking broader conditions critical for habitat modeling.

To this end, this paper proposes a habitat modeling framework that integrates Kernel Density Estimation (KDE)[23] with a semantic segmentation model, Segformer[24], by framing the habitat prediction problem as a pixel-wise classification task. Specifically, KDE is employed to expand the presence-only data into presence-absence data without prior expert knowledge, thereby reducing individual bias. The Segformer model enables the integration of multi-scale surrounding environmental effects into comprehensive species distribution prediction. To evaluate the performance of our proposed method, MaxEnt is selected as a baseline of traditional SDMs.

Methodology

Data preparation

For species distribution modeling using a semantic segmentation-based deep learning approach, environmental variables formatted as rasters are utilized as model inputs. Correspondingly, presence/absence labels, also in raster format, function as the output labels for model training.

Pre-processing of environmental data

First, the environmental variables, including bioclimatic variables, elevation, and water area proportion, are resampled to the same spatial resolution. Second, they are treated as individual bands and combined into a multiband image (i.e., a stack of raster layers). Considering the time-lag between environmental change and species habitat migration, when choosing the environmental data, the recording time of the data is carefully considered to ensure alignment.

Presence-absence labeling

KDE is used to create pseudo-absence observation data in raster format without expert knowledge support and individual bias (Fig. 1), which addresses the lack of negative data of species observations.

KDE creates a smooth surface for each presence point, and the height of the surface at each location depends on the distance to the presence point, with the surface being higher at locations closer to the data point. A search radius parameter is used to define the extent of the surface, with the surface height being set to zero outside this radius. Repeating this process for all data points generates overlapping surfaces, which are then combined to form a continuous surface map of the entire area. Finally, the continuous surfaces are converted to a raster grid, where the value of each pixel represents the observation density. The density of the final result is calculated by:

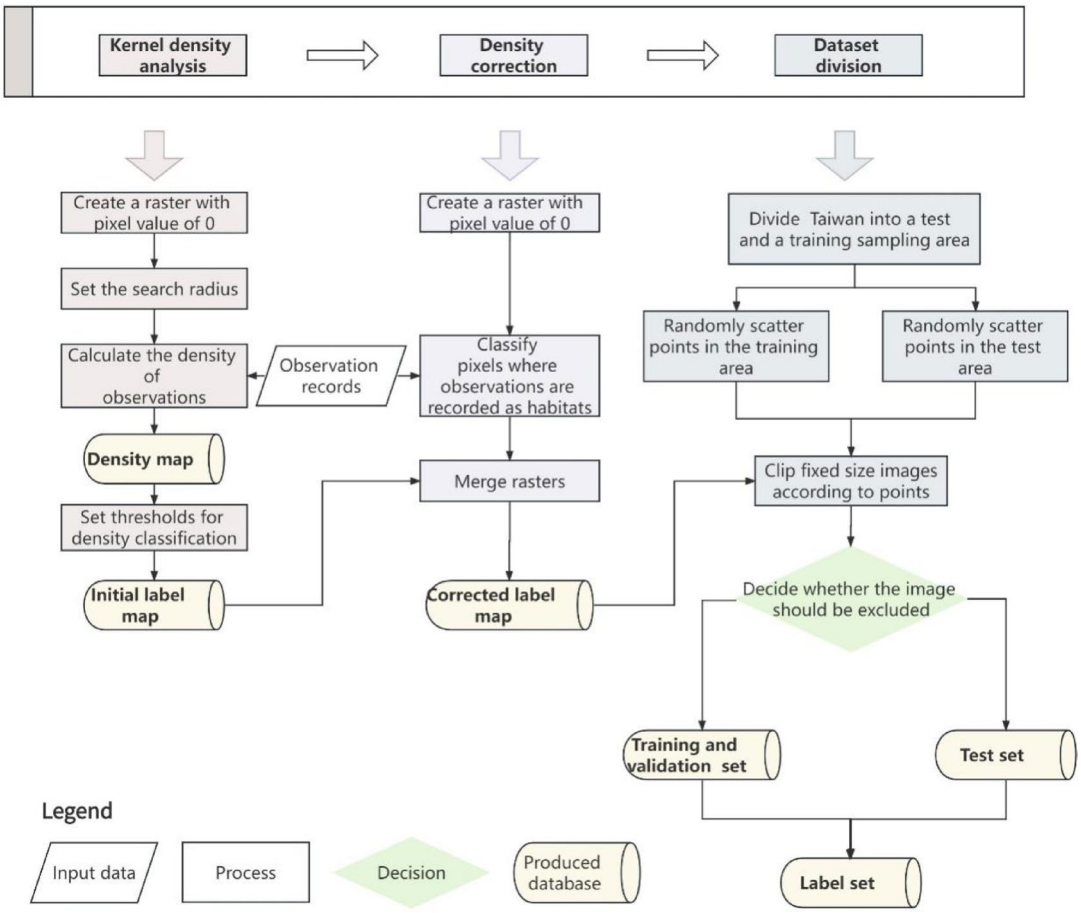

Fig. 1. The detailed workflow for the creation of the presence-absence label set.

$$\mathrm{h}(\mathrm{x},\mathrm{y}) = \frac{1}{r^2}\sum{}_{i=1}^{n}\left[\frac{3}{\pi}\left(1-\left(\frac{d_i}{r}\right)^2\right)^2\right]^2 \tag{1}$$

where $\mathrm{h}(\mathrm{x},\mathrm{y})$ is the height of the surface which represents the pixel value, r is the search radius, i is the number of the input point and d_i is the distance of the i-th point to that (x,y) pixel.

Based on the calculated kernel density values, pixels are classified from low to high density value into: non-habitat, potential non-habitat, potential habitat, and habitat. Pixels with a kernel density of 0 are classified as non-habitat. For the remaining pixels two thresholds are applied. Pixels with a density value below the mean value are classified as potential non-habitat. Pixels with a density value between the mean value and the mean plus triple the standard deviation are classified as potential habitat, and pixels with a density value above the mean plus triple the standard deviation are classified as habitat.

It is typically assumed that locations with observation records are habitats. However, the kernel density formula reveals that even with clear presence records, locations far from other observations might yield low kernel density values. This could lead to pixels containing presence records being classified as non-habitat. To correct for this bias, a correction process is necessary. After correction, all pixels with observation records are classified as habitat, and the original categories of other pixels are retained according to the selected thresholds. The final result is a four-category map with a spatial resolution of 50 m.

Dataset preparation

The semantic segmentation model divides an image into different parts, and then each part can be assigned to the corresponding category. Given that the input for such models must be fixed-size raster layers, it's crucial to clip both the multiband image of environmental variables and the pre-processed label image. The appropriate input size for the model should be carefully chosen based on the species' activity radius.

Supervised deep learning methods usually require dividing the database into three subsets for model training: the training set, the validation set, and the test set. The test set is not engaged in the training process, and its information must remain unknown to the model. Therefore, the following steps are designed carefully, as illustrated in Fig. 2:

1. Divide the study area into several equal parts according to latitude;

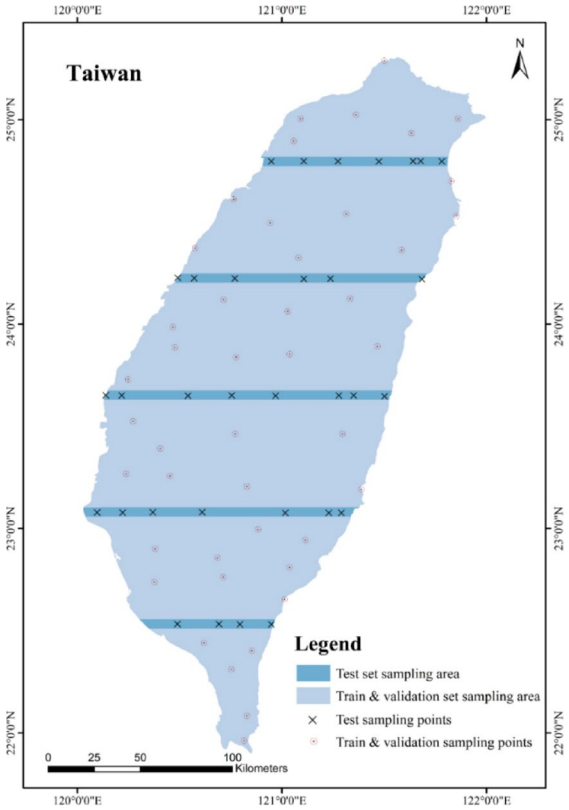

Fig. 2. The division of the training and validation sampling area and the test sampling area. The points on the map are demonstratively located.

2. Each part is split to create a strip zone;
3. These strip zones are considered the test set sampling area, with the remaining regions allocated for training and validation set sampling;
4. A certain number of points, which are the centroids of the fixed-size window, are scattered randomly within the study area;
5. Based on the centroids, combined environmental and label rasters are clipped and numbered sequentially;
6. To ensure a balanced representation of each pixel class, pairs containing more than 80% of pixels from a single class are excluded.

Segformer

Segformer, as one of the latest semantic segmentation models, combining a mixed-transformer encoder and a lightweight multilayer perceptron (MLP), is a lightweight, efficient neural network (Fig. 3)[24]. The mixed-transformer encoder (MiT) is a modification of the vision transformer by mixing the same architecture with different sizes.

Hierarchical transformer encoder

The self-attention is a specific type of attention mechanism and was initially proposed to solve the problem that RNNs are not able to directly model long-range dependencies[25,26].

The algorithm of the self-attention mechanism can be divided into three steps (Fig. 4):

(1) the query vector and key-value is computed based on input,

$$q = W_q x \tag{2}$$

$$k = W_k x \tag{3}$$

where W_q and W_k is the weights of query and key, respectively;

(2) the attention coefficients are produced based on the query vector and key-value,

$$att = \mathrm{softmax}\left(\frac{q \times k^T}{\sqrt{d}}\right) \tag{4}$$

where d is the dimension of the feature;

and (3) the result is calculated by multiplying the attention coefficients by the weights of the feature values

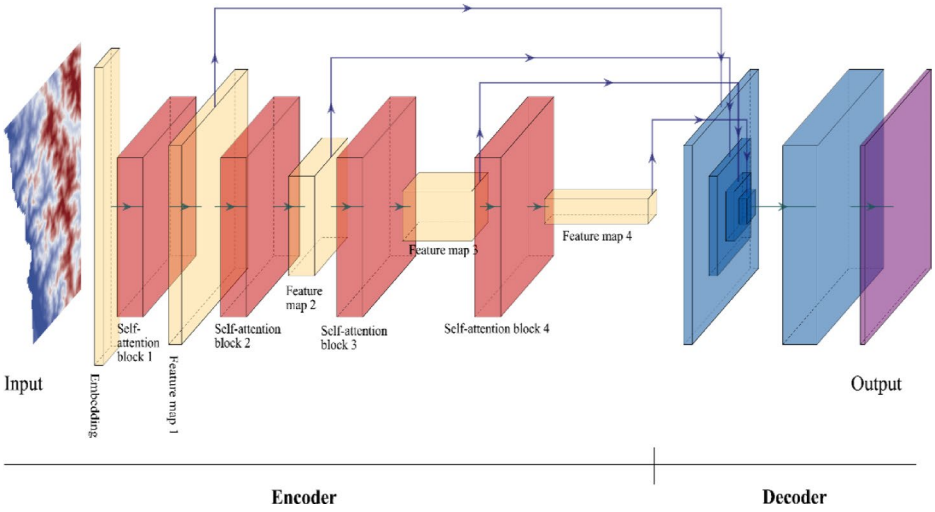

Fig. 3. The framework of Segformer network.

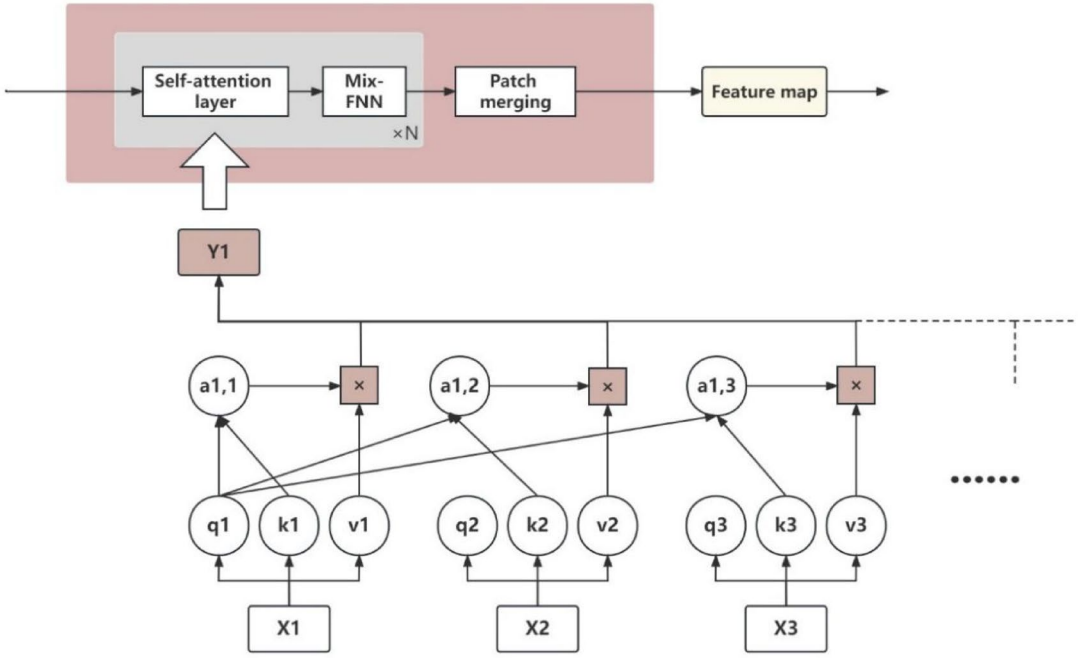

Fig. 4. The workflow of self-attention mechanism, ×N means there are N heads.

$$v = W_v x \tag{5}$$

$$result = \sum att \times v \tag{6}$$

where W_v is the weights of the feature value.

A single self-attention model may overly focus on the current position when encoding information, leading to the problem of being overly concentrated at a certain position[27]. The diverse information may be merged as knowledge, such as capturing dependencies of different ranges within the sequence, ultimately leading to a full feature description, by applying the self-attention method multiple times to learn varied results.

Segformer extracts four different resolution feature maps by downsampling to 1/4 of the original size each time. This constructs the hierarchical structure. The detailed to coarse feature maps extracted by 4-stage self-attention blocks provide accurate information efficiently. This further ensures the efficiency and accuracy of the model.

Mix feedforward neural network

Since semantic segmentation can be treated as a pixel-level classification task, the positional information between patches is not critical to the result, and it is not necessary to design explicit position encoding rules like the original transformer. Therefore, Enze Xie et al. designed a feedforward neural network (FNN) incorporating convolution, named Mix-FNN[24]. The Mix-FNN works as follows:

$$x_{out} = MLP\left(GELU\left(Conv_{3\times 3}\left(MLP\left(x_{in}\right)\right)\right)\right) + x_{in} \tag{7}$$

where, x_{in} is the feature provided by self-attention layers, $GELU$ is the activation function.

Gaussian error linear unit (GELU), incorporating the idea of stochastic regularity, are a probabilistic description of the input. The equation can be approximated as

$$x_{actv} = \frac{1}{2}x(1+\tanh\left(\sqrt{\frac{2}{\pi}}\left(x+0.044715x^3\right)\right)) \tag{8}$$

MLP decoder

MLP is one of the earliest models of artificial intelligence, which includes several fully connected layers and activation functions[28]. The fully connected layer is usually followed by an activation function to achieve a non-linear transformation. The fully-connected layers include two variables, weights and bias,

$$X_{out} = W_{m\times n}X_{in} + b_m \tag{9}$$

where $W_{m\times n}$ and b_m are the weight matrix and the bias, respectively, and n and m are the dimensions of the input and the output, respectively.

In Segformer, Rectified Linear Unit (ReLU) is selected as the activation function for the lightweight MLP decoder,

$$X_{actv} = \max(0, X_{out}) \tag{10}$$

where X_{out} is the output of the fully connected layer.

The ReLU ensures that the output is greater than or equal to 0 (Fig. 5), which is consistent with the range of label values we produce, i.e., four natural numbers representing four categories of habitat status.

The Segformer decoding workflow is divided into four steps (Fig. 6). Firstly, the feature maps at different spatial resolutions are decoded by using MLP. Secondly, the decoding results are resampled by a bilinear algorithm to the same resolution as the input image. Thirdly, the four resampled tensors with uniform resolution are merged into one tensor. Another MLP decodes the tensor to convert the features to the target class.

Loss function

Since the label set has four categories, the semantic segmentation task can be regarded as a multiclass classification problem. The cross-entropy function[29] is selected in this case. Considering the uneven proportion of categories in the labels, the cross-entropy loss function is computed with the category weights:

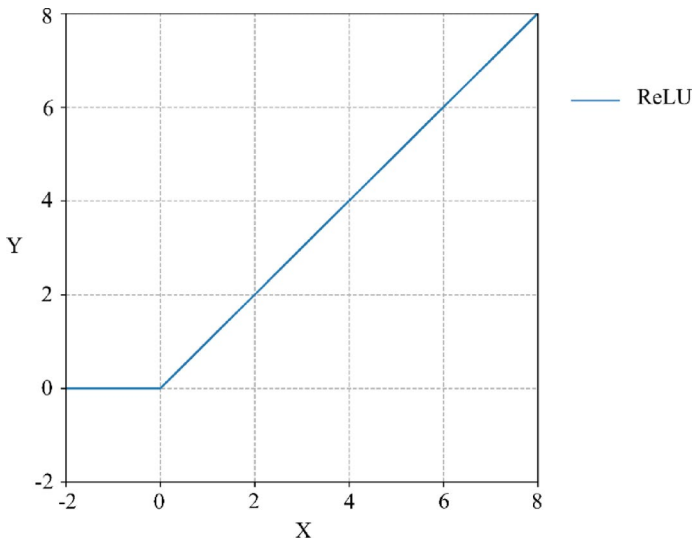

Fig. 5. Rectified linear unit (ReLU) function.

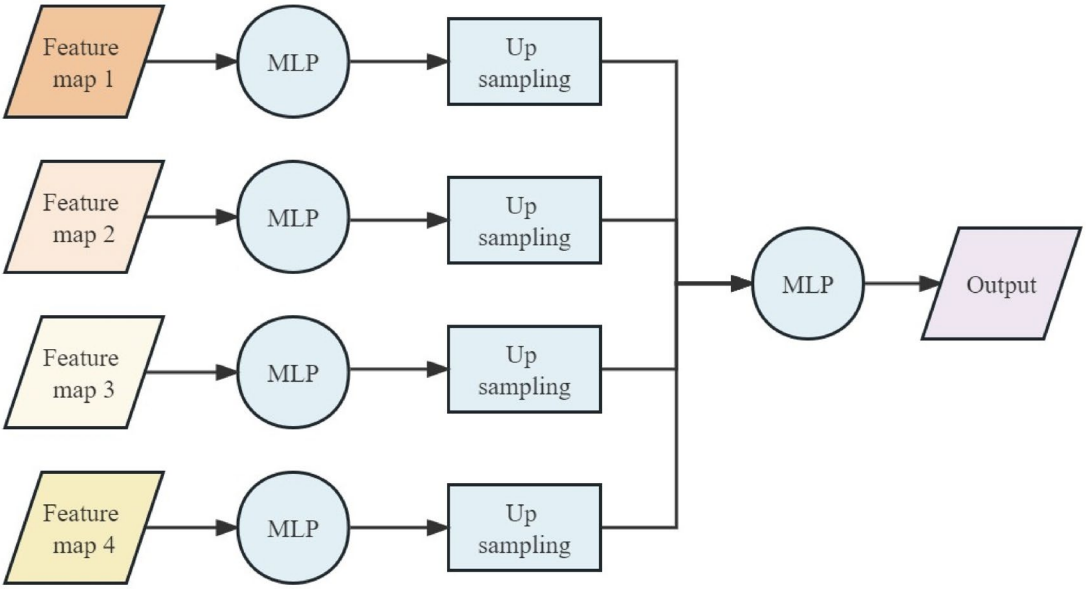

Fig. 6. The workflow of MLP decoder.

$$w_i = \frac{1}{\frac{n_i}{\sum_i^I n_i}} \tag{11}$$

where n_i represents the pixel number of the ith category.

The cross-entropy loss is calculated as follows,

$$\text{loss} = -\frac{1}{J}\sum_j^J \sum_i^I w_i f_{j,i} \log p_{j,i} \tag{12}$$

where, J is the total number of samples, I is the total number of categories classified, $p_{j,i}$ is the probability of the softmax output, i.e., the probability that the jth sample is predicted to result in class i, and $f_{j,i}$ is the indicator function with a value of 1 when the jth sample is correctly classified and 0 otherwise.

Verification of segformer

MaxEnt

Information entropy is an indicator used to express the degree of uncertainty. The higher the uncertainty, the greater the entropy:

$$\mathrm{H}(\mathrm{X}) = -\sum_i p_i \log p_i \tag{13}$$

where X is the data set and p_i is the probability of occurrence of the i th element in X.

In MaxEnt, it is believed that the model with the highest entropy is the most suitable choice. Two important constraints are implied: (1) if a specific condition is known, the total amount of the probability distribution of the result should be 1; (2) the expectation of the probability distribution of the extracted feature function concerning the training data set is equal to the expectation of the distribution for the model. The conditional entropy is calculated as follows:

$$\mathrm{H}(\mathrm{P}) = \sum_x P(x)\, H(Y|X=x) = -\sum_{x,y} \overline{P}(x)\, P(y|x) \log P(y|x) \tag{14}$$

The maximum entropy can be represented as

$$\max_P \mathrm{H}(\mathrm{P}) = -\sum_{x,y} \overline{P}(x)\, P(y|x) \log P(y|x) \tag{15}$$

i.e.,

$$\min_P -\mathrm{H}(\mathrm{P}) = \sum_{x,y} \overline{P}(x)\, P(y|x) \log P(y|x) \tag{16}$$

The constrained optimization problems are usually solved by introducing Lagrange multipliers.

The MaxEnt software provides the capability to automatically fit species observation data and environmental variables in ASCII format to model the relationships between the pairs of environment and species observation data (https://biodiversityinformatics.amnh.org/open_source/maxent/)[30]. In this study, MaxEnt is chosen as the representative of traditional methods, and is compared with the selected semantic segmentation model, which can confirm the feasibility of the segmentation model. The input data of MaxEnt in the feasibility experiment are raw observation points and environmental variables with a spatial resolution of 30 s. A random 80% subset of observation points is used for training, with the remaining 20% allocated to the validation set.

Model evaluation

The indexes commonly used for the evaluation of the semantic segmentation methods chosen for this study are Pixel Accuracy (PA) and Intersection-over-Union (IoU). PA is the percentage of pixels in the image that are correctly classified[31], i.e., the number of correctly classified pixels is divided by the total amount of pixels,

$$\mathrm{PA} = \frac{\sum_{i=0}^{I} p_{ii}}{\sum_{i=0}^{I} \sum_{j=0}^{I} p_{ij}} \tag{17}$$

where I represents how many categories are involved, p_{ij} represents the number of pixels that are class i but predicted as j.

PA can also be represented and calculated by the confusion matrix (Table 1). Therefore, PA can also be expressed as

$$\mathrm{PA} = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

The IoU is the area of overlap between the prediction result and the labels divided by the joint area between them[31], i.e., the intersection of the prediction and the label is divided by the union of the two,

$$\mathrm{IoU} = \frac{|A \bigcap B|}{|A \bigcup B|} = \frac{TP}{TP + FP + FN} \tag{19}$$

For each class, there is an IoU. Therefore, Segformer has four IoUs. To evaluate the method comprehensively, the weighted average of IoUs is calculated and used in the study.

Traditional species distribution models, like MaxEnt, use the area under the receiver operating characteristic (ROC) curve as a performance index. The ROC curve, also known as the sensitivity curve, is a smooth curve plotted with the horizontal and vertical axes being the false positive rate and the true positive rate, respectively, by constantly adjusting the classification threshold to obtain multiple pairs of FP rates and TP rates. AUC is the area sandwiched between the ROC curve and the horizontal axis[32]. However, in our case, Segformer has only one pair of FP rate and TP rate. The "curve" of Segformer is degenerated into a polyline connecting three points, i.e., the origin, (FP rate, TP rate) point and (1,1) point. The area sandwiched between the polyline to the horizontal axis acts as an approximate AUC.

Experiment and results

Study area

With the Pacific Ocean to the east and the Taiwan Strait to the west, Taiwan is situated on the continental shelf in the southeast of China (Fig. 7). The diverse geographical conditions and climate characteristics of Taiwan provide a suitable environment for various species, resulting in abundant natural and biological resources. In the bird checklist[33,34], a total of 674 bird species were recorded in Taiwan, including 171 species of vagrants, 29 species of seabirds, 176 species of migratory birds, and 29 endemic species.

Dataset

Observatory dataset

eBird is a bird observation platform sponsored by the Cornell Lab of Ornithology and the National Audubon Society. It gathers information from the activities of all birdwatchers in the world by using standardized data-gathering techniques to comprehensively document the distribution of birds. Two bird families are chosen for the experiments in this paper, which have a large number of observation records in Taiwan and fulfill the requirements of the method. The time window of observatory records is from 2000 to 2023.

Sandpipers *(Scolopacidae)* belonging to the class Aves and order Charadriiformes, consist of 23 genera and 83 species of wading birds of medium to small size. They are easily found in shallow waters such as swamps,

	Prediction	
Label	**Positive**	**Negative**
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

Table 1. The confusion matrix for the classification task.

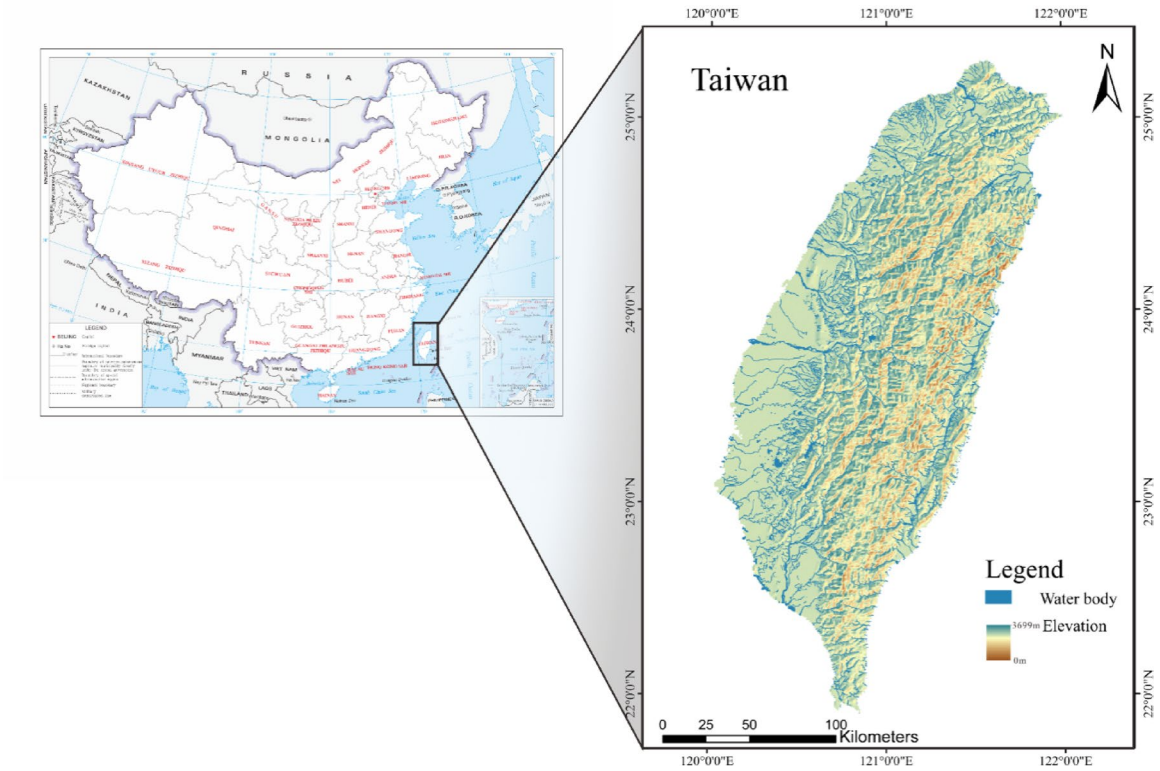

Fig. 7. The study area, the main island of Taiwan. The map (left) is downloaded from Standard Map Service System (http://bzdt.ch.mnr.gov.cn/) and the map (right) is generated by the author using ArcGIS 10.4 (https://www.esri.com).

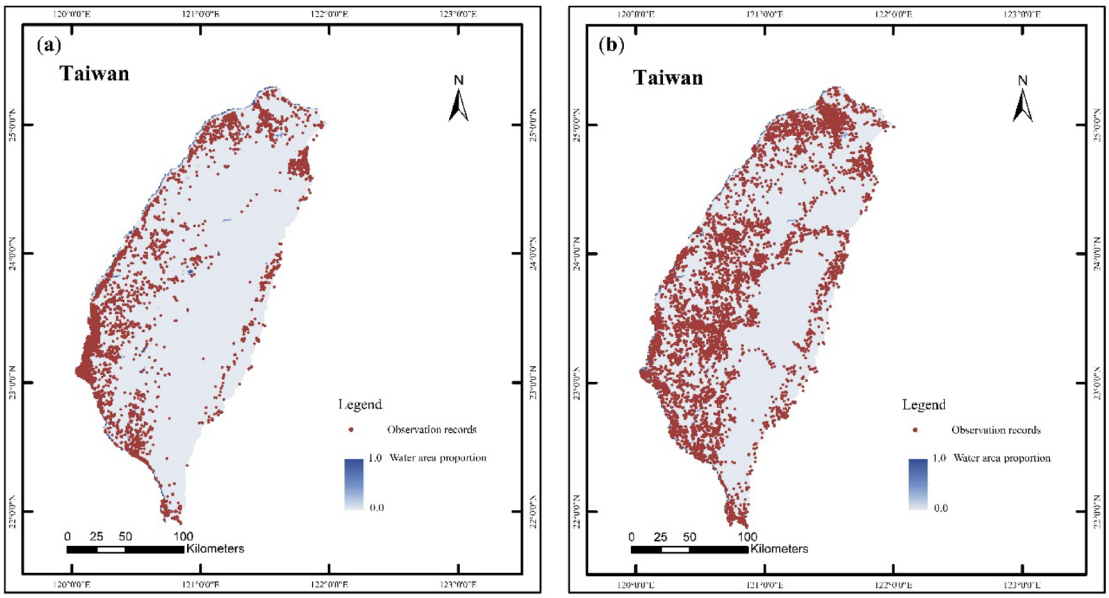

Fig. 8. The point map of the observation records of (**a**) the Sandpiper family and (**b**) the Swallow family on the main island of Taiwan. The maps are generated by the author using ArcGIS 10.4 (https://www.esri.com).

lakes, coastlines, and rivers. The birds belonging to the Sandpiper family share a similar set of characteristics, behaviors, and environmental preferences, making them a suitable bird family for the study of relationships between environment and their presence. A total of 407,072 observations of Sandpipers in Taiwan were used in this study (Fig. 8), including 15 categories such as the commonly found Calidris, Tringa, and Actitis genera.

The Swallow *(Hirundinidae)* family, which belongs to the order Passeriformes and suborder Passeri, comprises 88 species in 19 genera. The Swallow family has similar habits and is one of the well-known bird families with a close relationship to humans[35]. They inhabit mudflats, ponds, lakes, rivers, and cliffs in mountainous regions and can be found in a wide range, including grasslands, open woodlands, savannas, marshes, mangrove forests, and scrublands, from sea level to high altitudes[36]. Due to their beneficial role as insectivores, they are allowed to nest in urban and agricultural landscapes. In the case study, a total of 355,396 valid records of Swallows are obtained from Taiwan (Fig. 8), including only four genera, Hirundo (72%), Cecropis (15%), Riparia (10%), and Delichon (3%).

To account for the lagging effect of environmental conditions on species distribution[37], environmental variables collected before 2000 and bird observation records from 2000 to March 2023 were used in the experiment.

Environmental dataset

Three categories of environmental variables data are considered in the study: bioclimatic variables, elevation, and water area proportion (Fig. 9).

Researchers have found that the growth and breeding of species rely heavily on environmental variables such as suitable temperature and precipitation[37–39]. The dataset of WorldClim (https://www.worldclim.org/data/bioclim.html) selects 19 variables (Table 2), focusing on changes in temperature and precipitation and providing an efficient representation of the biologically relevant climate of the study area[40]. The bioclimatic dataset from WorldClim, at a spatial resolution of 30 arc-seconds (approximately 0.93 km), represents long-term climate averages for the period 1970–2000.

Elevation directly reflects the changes of the terrain. Besides the terrain difference, elevation data not only provides direct information about altitude but also includes information about variation in temperature, soil type, and vegetation distribution[41,42]. Elevation data, which includes information on vegetation distribution and topography, is an important environmental variable for mapping habitats. WorldClim also provides the organized global elevation data with a spatial resolution of 30 s (≈0.93 km) from Shuttle Radar Topography Mission. The data was measured during February 11–22, 2000.

Water provides a vital guarantee for the sustenance of life[43]. Apart from natural precipitation, water bodies such as rivers, lakes, and ponds in the natural environment serve as important water sources for avian survival. GPWv411: Water Area[44], published by Columbia University, provides the global estimated water area. This data grid includes an estimated proportion of water surface area per pixel at 1 km spatial resolution, and the data involved in this study is the average for the year 2000.

The above 21 images need to be resampled to the resolution of 50 m based on the cubic convolution interpolation method. Finally, the 21 images are integrated into stack of 21 raster layers.

Results of the sandpiper family

The habitat prediction map of the Sandpiper family in Taiwan is shown in Fig. 11 (left). The distribution pattern is concentrated. The predicted habitat is concentrated on the western coast and northeastern part of the island, with non-habitat concentrated in the inland of the island. The predicted habitat area is approximately 4848 km^2,

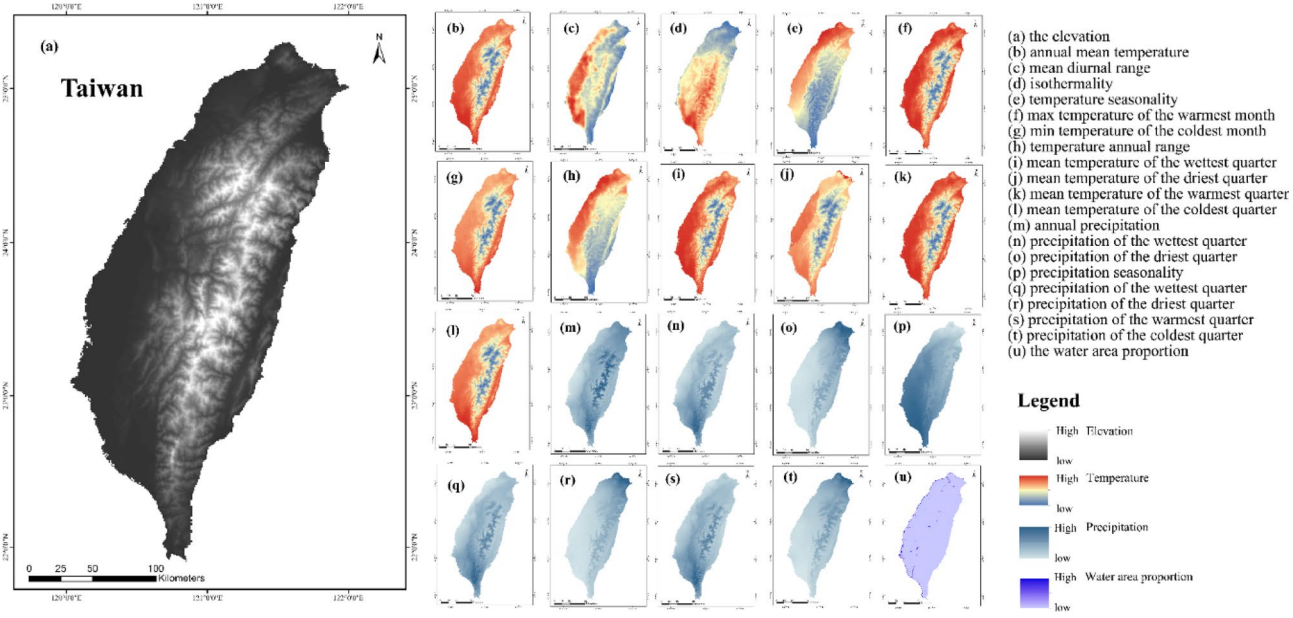

Fig. 9. 21 Environmental variables of Taiwan. The map is generated by the author using ArcGIS 10.4 (https://www.esri.com).

Nr.	Bioclimate index	Description
1	Annual mean temperature	The average of the weekly temperature records over a year
2	Mean diurnal range	The average of the difference of weekly maximum and minimum temperature over a year
3	Isothermality	Mean diurnal range divided by annual temperature range
4	Temperature seasonality	The temperature coefficient of variation
5	Max temperature of the warmest quarter	The maximum weekly temperature over the year
6	Min temperature of the coldest quarter	The minimum weekly temperature over the year
7	Temperature annual range	The difference of the max temperature of the warmest period and the min temperature of the coldest period
8	Mean temperature of the wettest quarter	The average of the weekly temperature in the wettest months
9	Mean temperature of the driest quarter	The average of the weekly temperature in the driest months
10	Mean temperature of the warmest quarter	The average of the weekly temperature in the warmest months
11	Mean temperature of the coldest quarter	The average of the weekly temperature in the coldest months
12	Annual precipitation	The sum of the monthly precipitation over a year
13	Precipitation of the wettest quarter	The precipitation sum of the wettest month
14	Precipitation of the driest quarter	The precipitation sum of the driest month
15	Precipitation seasonality	The precipitation coefficient of variation
16	Precipitation of the wettest quarter	The precipitation sum of the wettest period
17	Precipitation of the driest quarter	The precipitation sum of the driest period
18	Precipitation of the warmest quarter	The precipitation sum of the warmest period
19	Precipitation of the coldest quarter	The precipitation sum of the coldest period

Table 2. The details of the bioclimate variables chosen by the worldclim website.

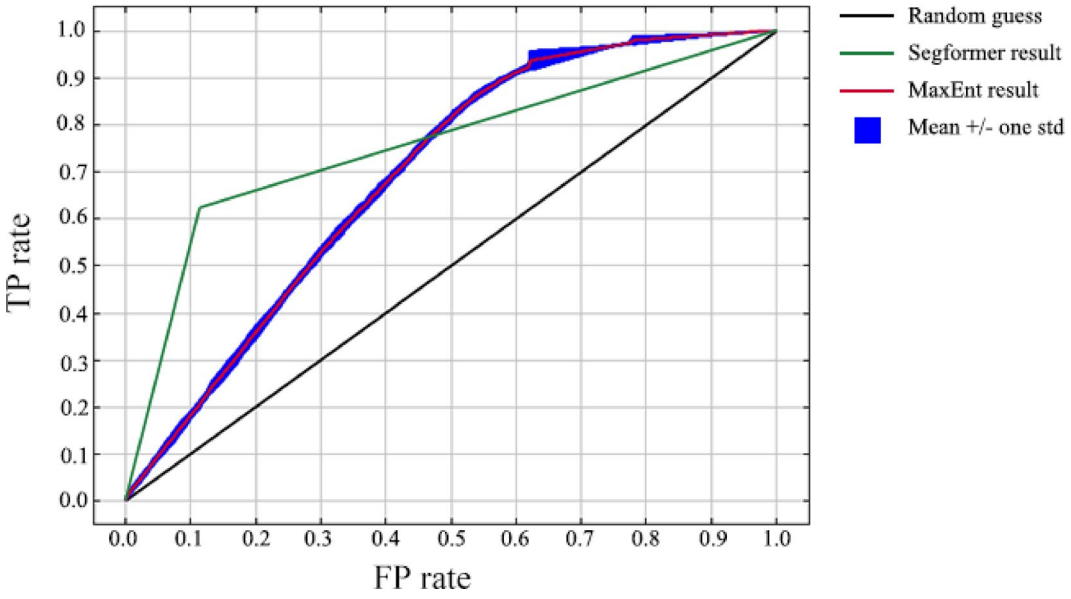

Fig. 10. The model evaluation results of Segformer and MaxEnt.

accounting for 11.3% of the island's area. 274,513 of the 327,934, observation records fall within pixels classified as habitat, which means 83.6% are correctly classified.

The average PA of Segformer is 0.65, the IoU is 0.54 and the approximate AUC is 0.76. The AUC of MaxEnt is the average AUC of 10 modelling runs, and the result AUC is 0.69 (Fig. 10). The performance of Segformer for the Sandpipers' winter habitat prediction is improved by 0.07 by integrating the spatial features of environmental variables.

As shown in Fig. 11, the predictions from Segformer and MaxEnt exhibit a similar pattern, with habitat suitability concentrated along coastal areas and flat regions. However, the habitat predicted by MaxEnt is more greatly influenced by the elevation variable, where the boundary of the non-habitat part overlaps to some extent with the 1000 m contour line. Although MaxEnt produces more detailed habitat maps than Segformer, its lower predictive performance and larger uncertain areas (highlighted in yellow) suggest overfitting. These uncertain regions offer little value for classification, as they provide no clear information.

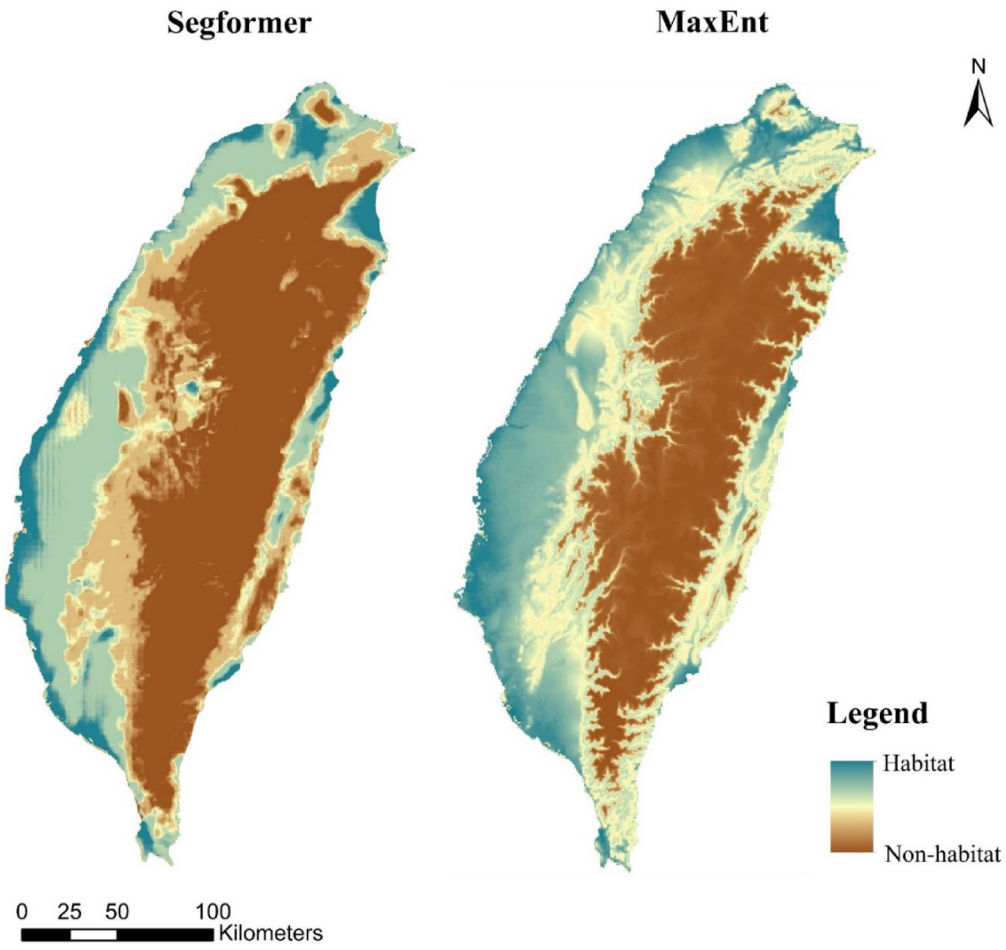

Fig. 11. The comparison of prediction results of Segformer and MaxEnt. The map is generated by the author using ArcGIS 10.4 (https://www.esri.com).

Results of the swallow family

The habitat prediction map of the Swallow family in Taiwan is shown in Fig. 12. The distribution pattern is wider than Sandpipers. The predicted habitat is concentrated on the western coast, with non-habitat concentrated in the inland of the island. The predicted habitat area is approximately 20,159 km^2, accounting for 46.98% of the island's area. 257,837 of the 355,396 observation records fall within pixels classified as habitat, which means 72.5% are correctly classified.

In terms of the habitat prediction results of the Swallow family in Taiwan, it can be seen that the prediction accuracy (PA 0.54) of Segformer for the Swallows' habitats is significantly lower than that for the Sandpiper family (PA 0.65), and the estimated AUC (0.68) is also slightly lower than the traditional method (0.69), Fig. 13. The application of Segformer in the habitat prediction of Swallows does not perform better than the traditional method.

Discussion

In the experiment, a total of 21 environmental factors were input without screening. Among them, the 21 variables can be categorized into four groups: temperature, precipitation, topography, and water resources. There are certain correlations among the 21 independent variables. According to the custom of traditional data analysis, the pre-processing needs to consider the correlation of independent variables and screen the data. Traditional data analysis often requires pre-processing to account for correlations between independent variables, with different model parameters needed for different cases. This can limit the generalizability of the method. In contrast, Segformer, a deep learning model, automatically extracts features from independent variables and establishes appropriate spatial configurations.

Kernel density analysis expands the original observation data into a raster image covering the study area, which increases the amount of data and compensates for the lack of absence data. Expert knowledge is generally valuable for producing absence data. It can be subjectively biased with the risk of propagation in the data processing, and expert knowledge is not always available. Mathematical formulas of kernel density analysis not only compensate for the lack of absence data but also avoid personal knowledge bias that can lead to data bias, ensuring the reliability of the database.

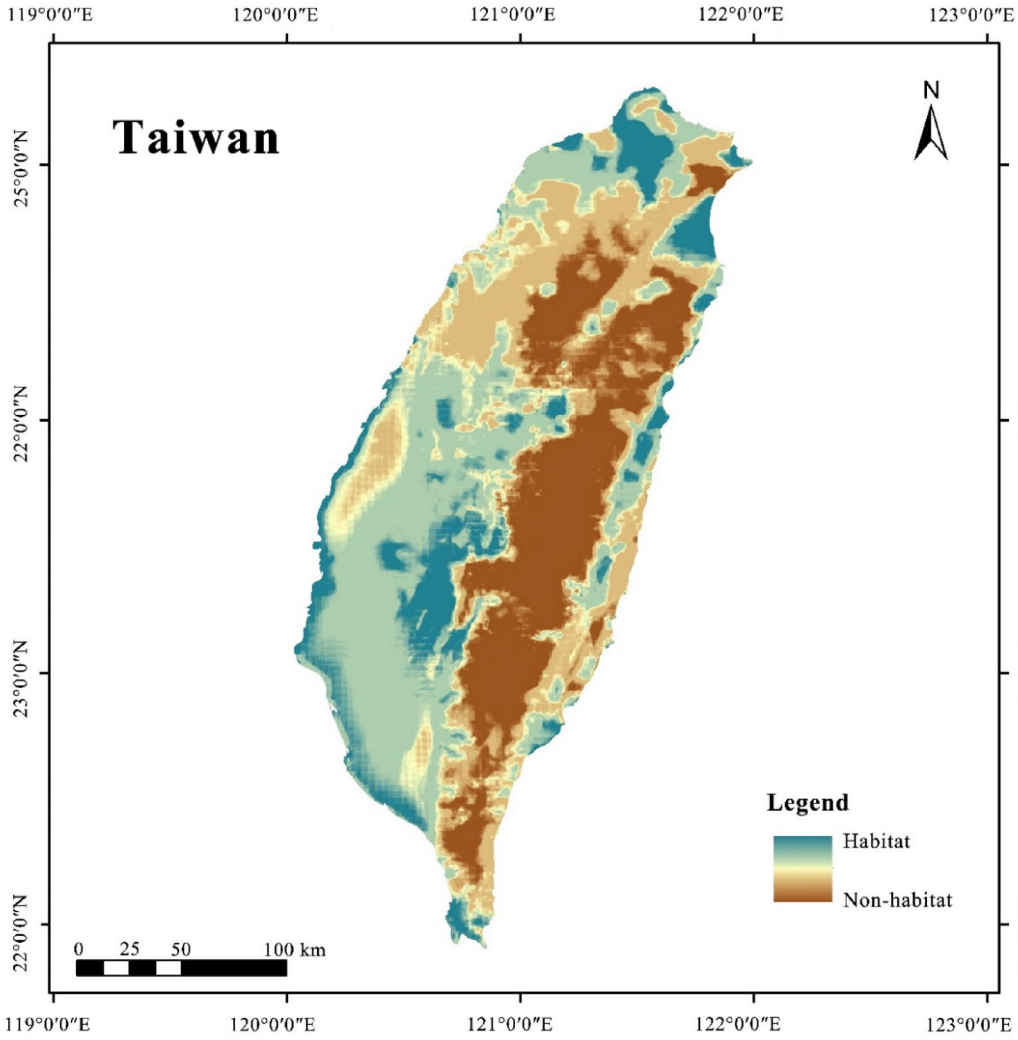

Fig. 12. The prediction of the Swallow family on the main island of Taiwan. The map is generated by the author using ArcGIS 10.4 (https://www.esri.com).

Segformer's advantages were not always apparent. In the results of the sandpiper habitat, Segformer is superior to the traditional method, while in the results of the swallow habitat, the performance of Segformer and the traditional method is on the same level. This implies that there are some limitations in the applicability of Segformer. We presume that the reasons for this contrast are as follows:

In the Segformer process, it is required to determine the size of the input raster. This size represents the environmental conditions of how many kilometers in a radius are considered to affect the habitat. Different bird species have different habitats and varying radii of activity. For example, if a bird has a radius of only 5 km, but all environments within a radius of 10 km are considered, it will introduce misleading information into the model, which will reduce the accuracy of the prediction. Studies have shown that Sandpipers have an activity range of approximately 64 km^2 during the breeding period and 118 km^2 during the non-breeding period, corresponding to activity radii of 4 km and 6 km, respectively[45]. Therefore, the area covered by the input raster of the semantic segmentation model should be slightly larger than the activity radii, which is 12 km×12 km for Sandpiper. In the experiment, the input raster is 12.8 km×12.8 km, which is consistent with the above conclusion.

However, the active radius of the Swallow family is usually no more than 500 m[46], while the size of the label and the stack of raster layers of the environment selected for Segformer is 12.8×12.8 km^2, which is 25 times greater than the activity range and introduce invalid information. It is essential to carefully consider the size of the input raster to ensure that it captures sufficient spatial context without introducing excessive or extraneous information that could hinder the model's predictive performance.

The approach of semantic segmentation classifies or recognizes targets based on features such as image texture and strong neighborhood differences. The Snipe family exhibits a clear preference for specific habitats, which is reflected as distinct spatial features across the 21 bands of environmental variables. This made it easier for the deep learning model to recognize and segment the Snipe's habitat. In contrast, the Swallow family has a broader distribution, with less pronounced environmental preferences. As a result, the spatial patterns in the

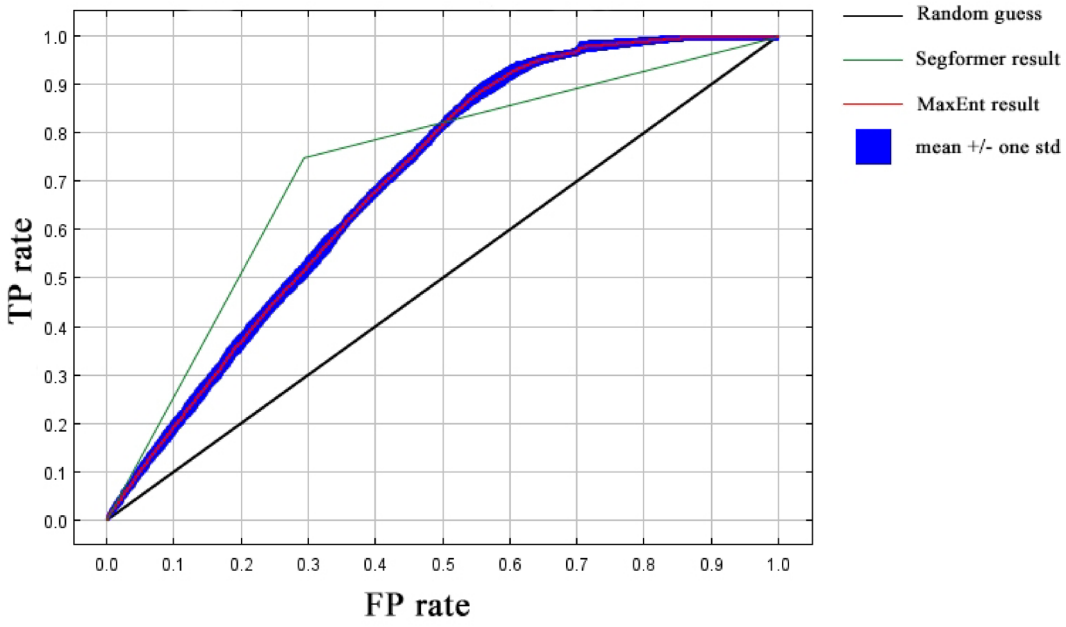

Fig. 13. The model evaluation results of Segformer and MaxEnt.

environmental variables are less distinct, making it more challenging for the model to identify and infer habitat features. Consequently, the results for the Swallow family were less accurate than expected.

In addition, the mapping result from Segformer has pixel-artefacts, which can be seen in Fig. 10. This is a common issue in image semantic segmentation, where the boundaries of the input can lead to poorer prediction performance. In this study, the model uses an input size of 256 × 256 pixels, and the study area is clipped into sections every 64 pixels. This ensures that each pixel is predicted within 16 stacked raster layers, and the final result is obtained by summing the predictions from these 16 layers. It is possible to use a more overlapping strategy, for example, 32 pixels apart, increasing the overlapping level to reduce the boundary effect even further. As mentioned above, Segformer has the properties of linear algebra, so the dimensions of the input are scalable, which means that the dimension of the input can be directly extended to the full range without clipping. However, the increase in dimension magnitude requires tremendous computational resources and system cache space.

Conclusion

This study proposes a Segformer-based method for habitat prediction, which consists of two main components. First, kernel density analysis is used to transform presence-only bird observation data into presence-absence data, enriching the dataset and providing labels for the semantic segmentation method. Second, Segformer is applied to predict bird habitats by extracting the relationships between environmental factors and species distribution. Compared to traditional species distribution models, our proposed method can more effectively integrate multi-scale environmental information, thereby capturing more complex spatial relationships between species and their habitats.

The prediction results for the Sandpiper family in Taiwan demonstrate that habitat prediction accuracy can be improved by incorporating surrounding environmental variables. However, the results for swallows indicate that the method performs less effectively for species with broader distributions. As a conclusion, Segformer can be seen as a complementary tool to traditional methods in specific cases, when the species being modeled has a large range of movement and relies on a broader surrounding environment.

Data availability

Data are already published and publicly available, with those items properly cited in this submission. Data sets used for this research are as follows: 1. eBird (https://doi.org/10.15468/aomfnb) is a bird observation platform sponsored by the Cornell Lab of Ornithology and the National Audubon Society (Auer T et al. 2022). 1.1 Dataset of the Sandpiper family was downloaded from eBird (https://www.gbif.org) using the following query: Species name [Scolopacidae], Year [2000 to 2023], Country or area [Chinese Taipei]0.1.2 Dataset of the Swallow family was downloaded from eBird (https://www.gbif.org) using the following query: Species name [Hirundinidae], Year [2000 to 2023], Country or area [Chinese Taipei]2. The dataset of WorldClim (https://doi.org/10.1002/joc.5086) selects 19 variables, focusing on changes in temperature and precipitation (Fick and Hijmans 2017)0.3. WorldClim (https://www.worldclim.org/) provides the organized global elevation data from Shuttle Radar Topography Mission (Fick and Hijmans 2017)0.4. Google Earth Engine dataset, GPWv411 (https://doi.org/10.7927/H49C6VHW): Water Area, provides the global estimated water area (Center for International Earth Science Information Network - CIESIN - Columbia University 2018).

Received: 4 January 2025; Accepted: 25 June 2025
Published online: 30 September 2025

References

1. Groom, M. J. et al. *Principles of Conservation Biology* (Sinauer Associates Sunderland, 2006).
2. Haddad, N. M. et al. Habitat fragmentation and its lasting impact on earth's ecosystems. *Sci. Adv.* **1**(2), e1500052 (2015).
3. Deneu, B. et al. Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLoS Comput. Biol.* **17**(4), e1008856 (2021).
4. Beery, S. et al. Species distribution modeling for machine learning practitioners: a review. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)*, 329–348. https://doi.org/10.1145/3460112.3471966 (2021).
5. Miller, J. Species distribution modeling. *Geogr. Compass*. **4**(6), 490–509 (2010).
6. Vasconcelos, R. P. et al. Predicting estuarine use patterns of juvenile fish with Generalized Linear Models. *Estuarine Coast. Shelf Sci.* **120**, 64–74 (2013).
7. Tom, Y. et al. Macrobenthic species response surfaces along estuarine gradients: prediction by logistic regression. *Mar. Ecol. Prog. Ser.* **225**, 79–95 (2002).
8. Lehmann, A. GIS modeling of submerged macrophyte distribution using generalized additive models. *Plant Ecol.* **139**(1), 113–124 (1998).
9. Leathwick, J. R., Elith, J. & Hastie, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model.* **199**(2), 188–196 (2006).
10. Miller, J. & Franklin, J. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecol. Model.* **157**(2), 227–247 (2002).
11. Moore, D. M., Lees, B. G. & Davey, S. M. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Environ. Manage.* **15**(1), 59–71 (1991).
12. Drake, J. M., Randin, C. & Guisan, A. Modelling ecological niches with support vector machines. *J. Appl. Ecol.* **43**(3), 424–432 (2006).
13. Yu, H., Cooper, A. R. & Infante, D. M. Improving species distribution model predictive accuracy using species abundance: application with boosted regression trees. *Ecol. Model.* **432**, 109202 (2020).
14. De'ath, G. & Fabricius, K. E. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*. **81**(11), 3178–3192 (2000).
15. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**(3–4), 231–259 (2006).
16. Stokland, J. N., Halvorsen, R. & Støa, B. Species distribution modelling—Effect of design and sample size of pseudo-absence observations. *Ecol. Model.* **222**(11), 1800–1809 (2011).
17. Iturbide, M. et al. A framework for species distribution modelling with improved pseudo-absence generation. *Ecol. Model.* **312**, 166–174 (2015).
18. Mcpherson, J. M., Jetz, W. & Rogers, D. J. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?. *J. Appl. Ecol.* **41**(5), 811–823 (2004).
19. Franklin, J. et al. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in Southern California. *Divers. Distrib.* **15**(1), 167–177 (2009).
20. Christin, S. et al. Applications for deep learning in ecology. *Methods Ecol. Evol.* **10**(10), 1632–1644 (2019).
21. Estopinan, J. et al. Deep species distribution modeling from sentinel-2 image time-series: A global scale analysis on the orchid family. *Front. Plant. Sci.* **13**, 839327 (2022).
22. Botella, C. et al. A deep learning approach to species distribution modelling. In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*, 169–199 (Springer International Publishing, 2018).
23. Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Routledge, 2018).
24. Xie, E. et al. SegFormer: simple and efficient design for semantic segmentation with transformers. *Adv. Neural. Inf. Process. Syst.* **34**, 12077–12090 (2021).
25. Ma, C. et al. Point-of-interest recommendation: exploiting self-attentive autoencoders with neighbor-aware influence. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy; Association for Computing Machinery.*, 697–706. https://doi.org/10.1145/3269206.3271733 (2018).
26. Wang, C. & Julian, M. Self-attentive sequential recommendation. In *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, F17–20 (2018).
27. Vaswani, A. et al. Attention is all you need. *Adv. Neural. Inf. Process. Syst.*, **30** (2017).
28. Gardner, M. W. & Dorling, S. R. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos. Environ.* **32**(14), 2627–2636 (1998).
29. Kline, D. M. & Berardi, V. L. Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Comput. Appl.* **14**, 310–318 (2005).
30. Phillips, S. J. et al. Opening the black box: an open-source release of maxent. *Ecography*. **40**(7), 887–893 (2017).
31. Wu, Z. et al. Segmentation of abnormal leaves of hydroponic lettuce based on DeepLabV3 + for robotic sorting. *Comput. Electron. Agric.* **190**, 106443 (2021).
32. Narkhede, S. Understanding auc-roc curve. *Towards Data Sci.* **26**(1), 220–227 (2018).
33. Hoyo, J. et al. *HBW and BirdLife International Illustrated Checklist of the Birds of the World*, Vol. 2 (2016).
34. Lin, D-L. & Pursner, S. *The State of Taiwan's Birds 2020*. (2020).
35. Turner, A. & Rose, C. *A Handbook To the Swallows and Martins of the World* (A&C Black, 2010).
36. Turner, A. Family Hirundinidae (swallows and martins). In *Handbook of the Birds of the World* Vol. **9**, 602–685 (2004).
37. Adler, P. B. & Levine, J. M. Contrasting relationships between precipitation and species richness in space and time. *Oikos* **116**(2), 221–232 (2007).
38. Evans, T. G., Diamond, S. E. & Kelly, M. W. Mechanistic species distribution modelling as a link between physiology and conservation. *Conserv. Physiol.* **3**(1), cov056 (2015).
39. Stenseth, N. C. & Mysterud, A. Climate, changing phenology, and other life history traits: Nonlinearity and match–mismatch to the environment. *Proc. Natl. Acad. Sci.* **99**(21), 13379–13381 (2002).
40. Xu, T. & Hutchinson, M. *Anuclim Version 6.1 User Guide* Vol. 90 (The Australian National University, Fenner School of Environment and Society, 2011).
41. Djukic, I. et al. Soil organic-matter stocks and characteristics along an alpine elevation gradient. *J. Plant Nutr. Soil Sci.* **173**(1), 30–38 (2010).
42. Whittaker, R. H. & Niering, W. A. Vegetation of the Santa Catalina mountains, Arizona. V. Biomass, production, and diversity along the elevation gradient. *Ecology* **56**(4), 771–790 (1975).
43. Hawkins, B. A. et al. Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**(12), 3105–3117 (2003).
44. Center For International Earth Science Information Network - CIESIN - Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11 [DS]. (2018).

45. Hill, J. M., Sandercock, B. K. & Renfrew, R. B. Migration patterns of upland sandpipers in the Western hemisphere. *Front. Ecol. Evol.* **7** (2019).
46. Ambrosini, R. et al. The distribution and colony size of barn swallows in relation to agricultural land use. *J. Appl. Ecol.* **39**(3), 524–534 (2002).

Author contributions

Lingjun Wang prepared the data, performed the experiments, visualized the results, wrote the main manuscript and revised it based on suggestions from other authors. Haofeng Tan participated in data preparation and main experiments, and provided computational facilities. Peng Luo provided useful suggestions on the experimental ideas and steps, and participated in the revision of the manuscript. Liqiu Meng provided constructive comments on the conceptual design and participated in the revision and correction of the manuscript . Fei Teng provided ideas and data support to the team and contributed greatly in the revision stage of the manuscript .

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.