Exploiting Time-Series Image-to-Image Translation to Expand the Range of Wildlife Habitat Analysis

Ruobing Zheng,[1,2] **Ze Luo,**[2] **Baoping Yan**[2]
[1]University of Chinese Academy of Sciences, Beijing 100049, China
[2]e-Science Technology and Application Laboratory,
Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
{zhengruobing, luoze, ybp}@cnic.cn

Abstract

Characterizing wildlife habitat is one of the main topics in animal ecology. Locational data obtained from radio tracking and field observation are widely used in habitat analysis. However, such sampling methods are costly and laborious, and insufficient relocations often prevent scientists from conducting large-range and long-term research. In this paper, we innovatively exploit the image-to-image translation technology to expand the range of wildlife habitat analysis. We proposed a novel approach for implementing time-series image-to-image translation via metric embedding. A siamese neural network is used to learn the Euclidean temporal embedding from the image space. This embedding produces temporal vectors which bring time information into the adversarial network. The well-trained framework could effectively map the probabilistic habitat models from remote sensing imagery, helping scientists get rid of the persistent dependence on animal relocations. We illustrate our approach in a real-world application for mapping the habitats of Bar-headed Geese at Qinghai Lake breeding ground. We compare our model against several baselines and achieve promising results.

Introduction

The analysis of space use and habitat selection by animals is a well-studied topic in ecology (Calenge 2006). The measurement of wildlife habitat is not only valuable in predicting the appearance of animals but also plays an important role in resource management and animal conservation. The original habitat studies used to define the space use of animals as a uniform area surrounded by borders (Mohr 1947). However, in reality, animals are unlikely to uniformly use their habitat. Therefore, some probabilistic habitat models (Worton 1989; Clark, Dunn, and Smith 1993) have emerged in the literature. These models take the form of a two-dimensional probability density map to represent the possibility of animals occurring, or score the habitat suitability at each location. We mainly focus on this type of habitat models in the following work.

Locational data are extensively used in characterizing wildlife habitat (Bayliss, Simonite, and Thompson 2005; Horne et al. 2007), either directly used in kernel density estimations or helping to capture the relationship between animals and their environment. In the old days, locational data were gathered by field observation or trapping methods. Such sampling ways are very laborious, which limited the sampling field and monitoring period. Nowadays, most locational data are automatically obtained by the radio-tracking device carried by animals. The development of GIS technology facilitates the sampling procedure, but the sample size is still far from enough. In practice, scientists often capture and mark a fixed amount of target species, receiving animal relocations during the validity period of the vulnerable GPS transmitters (Takekawa et al. 2010). The whole work is costly and in most cases a one time job. Therefore, the insufficient locational data still restrict the range of wildlife habitat analysis, both in the spatial and temporal dimension.

Habitat mappings (Nagendra et al. 2013), a type of ecological applications of remote sensing, provide alternatives for modeling habitat. They succeed in mapping the quality and extent of animal habitat from remote sensing imagery, which demonstrates the strong connection between wildlife and their environment. However, most of these studies can be hardly expected to map general habitat models. That is because they mostly focus on the specific species and map the self-defined habitat category or index (Lee et al. 2017). Moreover, their mapping models are mainly the traditional machine learning classifiers or regressors which work on individual pixels (Chegoonian, Mokhtarzade, and Valadan Zoej 2017). As we know, remote sensing imagery is highly structured. When mapping more complex target, such as a complete probability density map, the traditional pixel-based scheme is far from satisfactory (Zheng et al. 2018b).

To overcome this limitation, we pay attention to the emerging image-to-image translation technology. These studies aim to learn the mapping between input images and output images using a set of aligned image pairs, showing promise in various applications (Sangkloy et al. 2017; Pathak et al. 2016). Typically, pix2pix investigates Conditional Generative Adversarial Network (CGAN) as a general-purpose solution to image-to-image translation problems (Isola et al. 2016), leading to a substantial boost of sampling quality. Following studies further focus on some specific domain, such as multimodal (Zhu et al. 2017b) and unpaired (Zhu et al. 2017a) image-to-image translation.

Inspired by the above studies, we consider our mapping task as a specific image-to-image translation problem. Dif-

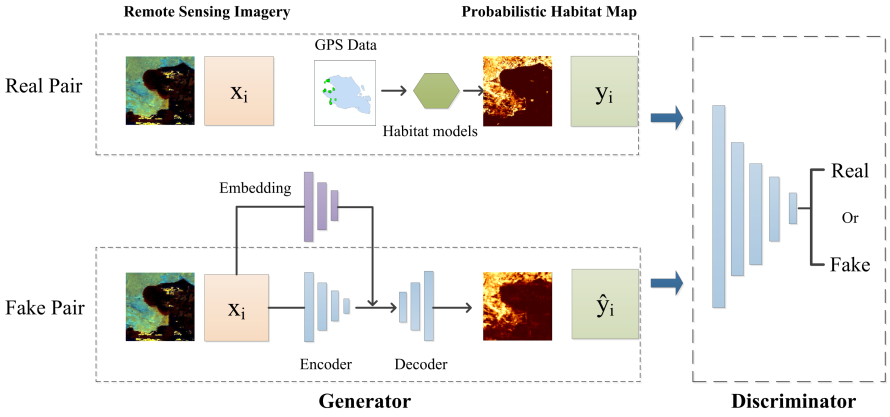

Figure 1: Using the time-series image-to-image translation framework to simulate probabilistic habitat models. The embedding network produces temporal vectors from the input images. The Generator learns to map the habitat map from remote sensing image. The Discriminator tries to classify the real and synthesized data-target pairs.

ferent from common applications, we focus on the mapping accuracy instead of the diversity in generated images. In this paper, we aim to learn a mapping from remote sensing imagery to the result of probabilistic habitat models in the supervised way. Our main contributions can be summarized as two part:

- We propose an innovative approach to simulate probabilistic habitat models from long-term support remote sensing imagery. We illustrate our approach in a real-world application for mapping the habitats of Bar-headed Geese at Qinghai Lake breeding ground. Our study enables scientists to expand the range of wildlife habitat analysis even the animal relocations are not enough for long-term and large-scale research.
- We propose a novel model to implement time-series image-to-image translation via metric embedding. A siamese neural network is used to learn the Euclidean temporal embedding from the image space. The embedding produces temporal vectors to replace the random noise in the generator. This strategy provides a new sight to bring temporal information into the adversarial network and shows promise in this specific application.

Proposed Model

We devise a time-series image-to-image translation framework to tackle our mapping problem. The framework consists of an adversarial network and a siamese network. The adversarial network is used to learn the mapping from a remote sensing image to the corresponding probability habitat map. The siamese network is used to learn a Euclidean temporal embedding from the image space. This embedding helps to produce temporal vectors that replace the random noise in the adversarial network, as shown in Figure 1.

Formulation

Adversarial Loss The CGAN (Mirza and Osindero 2014) enables scientists to use an image as the auxiliary information to control the generating process. The representative pix2pix (Isola et al. 2016) expands the generator's loss by adding the $\ell 1$ term balanced by λ. It learns a mapping from a type of images A and a random noise z, to another type of images B: $\{\mathrm{A,z}\} \rightarrow \mathrm{B}$. The objectives can be defined as:

$$\begin{aligned}\max_D V_{p2p}(D) =& \mathbb{E}_{\mathrm{A,B}\sim p(\mathrm{A,B})}[(\log\left(D(\mathrm{A,B})\right)]+ \\ & \mathbb{E}_{\mathrm{A}\sim p(\mathrm{A}),\mathrm{z}\sim p(\mathrm{z})}[\log\left(1-D(\mathrm{A},G(\mathrm{A},\mathrm{z}))\right)]\end{aligned} \tag{1}$$

$$\ell_1(G) = \mathbb{E}_{\mathrm{A,B}\sim p(\mathrm{A,B}),\mathrm{z}\sim p(\mathrm{z})}[\|\mathrm{B}-G(\mathrm{A},\mathrm{z})\|_1] \tag{2}$$

$$\begin{aligned}\min_G V_{p2p}(G) =& \mathbb{E}_{\mathrm{A}\sim p(\mathrm{A}),\mathrm{z}\sim p(\mathrm{z})}[\log\left(1-D(\mathrm{A},G(\mathrm{A},\mathrm{z}))\right)]+ \\ & \lambda\ell_1(G)\end{aligned} \tag{3}$$

In this letter, we aim to find the mapping: $X \rightarrow Y$, from a multi-band remote sensing image $X \in \mathbb{R}^{H\times W\times B}$ (B stands for the number of bands), to the probabilistic habitat map $Y \in \mathbb{R}^{H\times W\times 1}$. We apply two improvements to the original pix2pix model. First, we replace the sigmoid cross-entropy loss by a least-square loss for both generator and discriminator. The least-square loss has been reported to stabilize the training procedure and expedite the convergence (Mao et al. 2016). Second, we use the temporal vector $\phi(X)$ instead of the random noise z to involve time information. The $\phi(X)$ is produced by an embedding network ϕ which will be introduced later. The final objects can be written as:

$$\begin{aligned}\min_D V_F(D) =& \frac{1}{2}\mathbb{E}_{\mathrm{X,Y}\sim p(\mathrm{X,Y})}[(D(\mathrm{X},\mathrm{Y})-1)^2]+ \\ & \frac{1}{2}\mathbb{E}_{\mathrm{X}\sim p(\mathrm{X})}[D(\mathrm{X},G(\mathrm{X},\phi(X)))^2]\end{aligned} \tag{4}$$

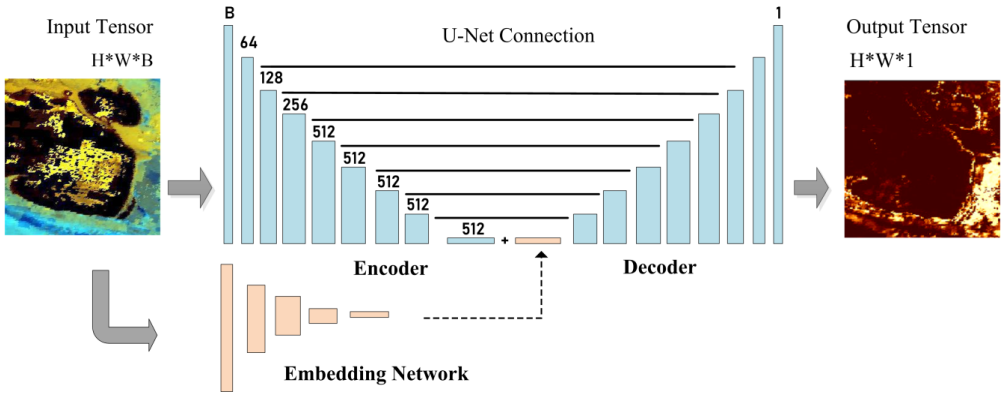

Figure 2: Architecture for the generator G in our time-series image-to-image translation framework. G generates a probabilistic habitat map from the input remote sensing image and the corresponding temporal vector.

$$\min_G V_F(G) = \frac{1}{2}\mathbb{E}_{\mathrm{X}\sim p(\mathrm{X})}[(D(\mathrm{X}, G(\mathrm{X}, \phi(X))) - 1)^2] + \lambda\mathbb{E}_{\mathrm{X},\mathrm{Y}\sim p(\mathrm{X},\mathrm{Y})}[\|\mathrm{Y} - G(\mathrm{X}, \phi(X))\|_1] \tag{5}$$

Euclidean Temporal Embedding Previous studies (Isola et al. 2016; Zhu et al. 2017b) report that the random noise z is used to add diversity for some one-to-many applications. But the mapping in this task is actually one-to-one, and our target is to improve the mapping accuracy. Besides, as the season changes, the environment may have the same appearance, but the habitat distribution changes a lot. So we need to involve the time information in the mapping model.

We create temporal vectors to replace the random noise z in the adversarial network. The temporal vector is produced by an embedding network ϕ that takes a remote sensing image as input and projects it into a Euclidean space. Inspired by a recent work (Courty, Flamary, and Ducoffe 2017), we use a siamese neural network (Bromley et al. 1994) to learn the embedding ϕ in the supervised way. This architecture was originally designed for metric learning and similarity learning (Chopra, Hadsell, and LeCun 2005). It is defined by replicating a network which takes as input a pair of samples, and learns a mapping to a new space with a contrastive loss. In this letter, we aim to force the distance between temporal vectors in the embedding space mimics the time difference between the original remote sensing images. For the remote sensing images (x_i^1 and x_i^2) in the same geolocation, we calculate the Euclidean distance y_i between their timestamps, and train the siamese network under the following objective:

$$V_E = \min_{\phi} \sum_i \left\| \left\| \phi(x_i^1) - \phi(x_i^2) \right\|^2 - y_i \right\|^2 \tag{6}$$

Network Architectures

Adversarial Network The adversarial network consists of a generative network G and a discriminative network D. We adapt the network architecture from those in (Isola et al. 2016) who have shown impressive results in different kinds of image-to-image translation tasks. The generator G is a deep convolutional encoder-decoder with "U-Net" (Ronneberger, Fischer, and Brox 2015) skip connection (Figure 2). The encoder extracts the high-level code from remote sensing bands while the decoder interprets and upsamples them to a full-size output. The discriminator D is a traditional convolutional classifier. It takes as input the combination of one remote sensing image and the corresponding habitat map, judging the image pair from real or synthetic. Both generator and discriminator use the combination as Convolution(Transpose)–BatchNorm–ReLU(Leaky) layers.

Siamese Network The siamese neural network contains two same neural networks that computes the embedding, as shown in Figure 3. The Euclidean temporal embedding ϕ is represented as such a encoder which has three convolutional layers with ReLu activations. The following linear layer produces the n-dimensional temporal vectors which defined by users. The complete encoder produces the temporal vector, $t = \phi(x)$, from the remote sensing space $\mathbb{R}^{H\times W\times B}$ to the embedded Euclidean space $\mathbb{R}^n$.

Training Strategy

We first train the siamese network to learn the temporal embedding from remote sensing images. To this end, we regroup the remote sensing images with the same geolocation. In each group, images have the same geographical structure but different timestamps. For each training step, we randomly select a pair of images from one group while computing the pairwise time differences. The siamese network learns an implicit embedding from these training pairs following the contrastive loss.

The next step is to train the adversarial network. Our training strategy is similar to most image-to-image translation frameworks where the input data is aligned image pairs. The difference is at the generator part. Both the encoder and em-

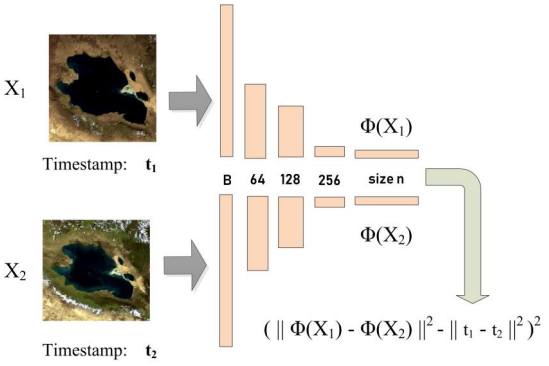

Figure 3: Architecture for the siamese neural network. Two embedding networks share the parameters and take a pair of remote sensing images as input. The Euclidean temporal embedding is learned by the contrastive loss.

bedding network take one remote sensing image as input. We concatenate their outputs and feed it into the decoder. Considering our dataset is smaller than common datasets, we employ data enlargement operations during the training. Mirroring, rotation and random jitter on input image pairs are used before each training steps.

Implementation

In this section, we implement our method in a real-world application for mapping the habitats of Bar-headed Geese in Qinghai Lake area during the breeding season.

Habitat Models

We test two types of habitat models respectively. Kernel UD (Worton 1989) stands for a type of Home Range(HR) estimators (Calenge 2011b) which only use locational data in habitat estimation. Kernel UD employs a Gaussian kernel on animals relocations to calculate the probability on each location within the defined area. Mahalanobis Habitat Suitability Index (HSI) (Clark, Dunn, and Smith 1993) represents another type of habitat models which utilize both locational data and environmental factors to score habitat suitability. It seeks the optimum in the animal-used ecological space from animal relocations and environmental factors and then measures the squared Mahalanobis distance from this optimum to each pixel (Calenge 2011a). The smaller distance for a given pixel means the better habitat suitability. Both of HR and HSI return a two-dimensional probability density map.

Study Area and Field Knowledge

Qinghai Lake is located in northeastern Qinghai province on the Tibetan-Qinghai Plateau. This lake covers the crossroads of several bird migration routes across Asia. We define a study area (96.6° and 102.4° E, and 34.2° and 38.8°N) where the lakes and surrounding wetlands serve as critical breeding sites for Bar-headed Goose, as shown in Figure 4. This special species gained global attention following the

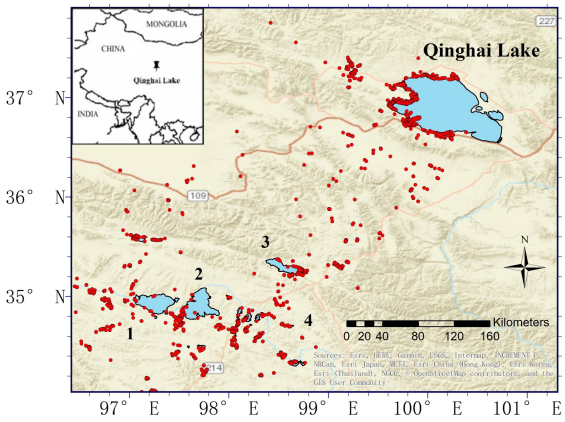

Figure 4: Location of Qinghai Lake and the map of the study area. The radio-tracking data of Bar-headed Geese are shown as red points. The key areas and codes: 1. Gyaring Lake, 2. Ngoring Lake, 3. Donggi Conag Lake, 4. Doucuo Co.

first large-scale outbreak of avian influenza at Qinghai Lake in the spring of 2005 (Chen et al. 2005). The habitat analysis of Bar-headed Goose could help to understand the disease transmission and guide the monitoring of disease outbreaks.

GPS and Remote Sensing Data

The GPS data were collected from the Bar-headed Geese which were captured and marked along Qinghai Lake on March 25-31, 2007. We attached a 45g solar-powered transmitter dorsally between the wings of each bird. After a preliminary review, we select five individuals with sufficient daily data points, for a total of 4449 samples during the breeding season in 2007 and 2008.

MODIS (Justice et al. 1998) Land Products (MOD09Q1 and MOD09A1 8-days L3) provide the basic reflectance bands in our experiments. Four environmental factors are considered based on the expert experience and field survey. We select Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) to determine the food availability (Dong et al. 2013; Zhang et al. 2018). The Normalized Difference Water Index (NDWI) helps us to estimate the access to water (Cappelle et al. 2010). When evaluating the shelter conditions, we refer to the MODIS land cover type following the previous work (Takekawa et al. 2010). For the HSI model, we use the environmental factors to compose input rasters. As for the HR model, we directly use the involved MODIS reflectance bands as input.

Preprocessing

We annotate the MODIS imagery with the probability maps produced from two selected habitat models using R package (Calenge 2006). To build the image-based data and target pair, we align each remote sensing image with the corresponding probability map on both the spatial and temporal resolution. At the temporal level, we subpackage GPS

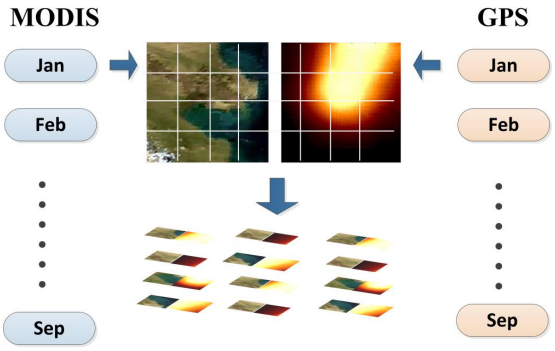

Figure 5: Illustration of the pre-processing procedure. We build the image-based data-target pairs to train the mapping framework.

data by every 8-days to match the time interval of MODIS imagery. To align each pixel, we produce the probability map using the remote sensing image as the background grid. All raster data are under the same projection (EPSG: 4326: WGS84) and resolution (250m). In the end, we slice big images into numerous 256×256 tiles and pair them, as shown in Figure 5.

Evaluation and Results

Baselines

We compare our model against several potential solutions to this mapping task, considering both habitat mapping and computer vision literature.

- **Decision Tree**: Decision tree (Rokach and Maimon 2008) is a traditional supervised model for both classification and regression problems. In habitat mapping literature, Classification and Regression Tree (CART) has shown promise in mapping the extent and quality of wildlife habitat (Pastick et al. 2015; Kobler and Adamic 2000). Here we examine it as a pixel-based baseline which stands for the traditional habitat mapping scheme.
- **CNN + ℓ_2 loss**: CNN with ℓ_2 loss is probably the most straightforward way to mapping continuous target in deep learning style. To exclude the impact from network architecture, we define the CNN as the same encoder-decoder structure as the proposed model.
- **pix2pix**: pix2pix is an established framework for various image-to-image translation applications. We examine this fundamental model as a baseline to verify whether the temporal embedding is an effective way to involve time information in the mapping model.

Metrics

- **Mapping Accuracy**: To quantitatively evaluate the continuous values in predicted probabilistic habitat maps, we employ Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to measure the mapping accuracy at the regression level.
- **Structural Similarity**: Since the probabilistic habitat map is also a structured image, we use the Structural Similarity Index (SSIM) (Wang et al. 2004) to measure the similarity between the mapping result and ground truth. SSIM is designed to improve the traditional regression metrics by taking the structure comparison into account. For consistency with RMSE and MAE, we measure the SSIM loss as $\mathcal{L}^{SSIM} = 1 - SSIM(p_i)$ for each pixel.

Qualitative Evaluation

As shown in Figure 6 and Figure 7, we observe that the pixel-based Decision Tree produces a large amount of noise in habitat maps, especially in mapping the HR model. The remaining image-based baselines eliminate the noise effectively, producing clearer and more recognizable results. However, CNN + ℓ_2 leads to the fuzzy outputs with larger high-probability zones than ground truth, both in HSI and HR. The pix2pix alleviates the blurring and produces sharper images for HSI. But when mapping HR, it brings some artifacts in the flat region while showing bias towards ground truth. Compared to pix2pix, our model reduce the artifacts and achieves more visually accurate results in mapping two types of habitat models.

Quantitative Evaluation

We quantitatively compare our model and baselines in three test strategies. We compare the overall performance by randomly selecting test samples from all image pairs. The rest pairs are used as the training and validation set. To compare the model performance in the spatial dimension, we manually select the samples in a specific location (Doucuo Co) as the test data and exclude the same area from the training set. In the time dimension, we test the samples only in a specific period (July 2008). In this scenario, we exclude the image pairs with the same timestamps from the training set to avoid temporal overlap.

Among the overall results in Table 1 and Table 2, we observe that Decision Tree obtains the low performance in mapping both HSI and HR on all three metrics. The high SSIM loss reveals that the scattered noise significantly reduces the structural similarity. Regarding two image-based baselines, both CNN + ℓ_2 and pix2pix have an obvious improvement than Decision Tree. Compared to CNN + ℓ_2, pix2pix leads to a better result in mapping HR but fails in simulating HSI. The blurring from CNN + ℓ_2 not only hampers the model from mapping a clear border of HR but also induces a high SSIM loss. The pix2pix show promise in constructing the HR. However, the diversity in generated results influences its performance in mapping HSI. Our model achieves the best performance compared all baselines on both regression metrics and structure similarity.

The other two test scenarios reveal more differences between the mapping models. At the spatial level, all image-based models take a step backward when facing new geomorphological structures, but Decision Tree is less affected. At the temporal level, benefited from the injection of temporal vectors, our model shows superiority than pix2pix and outperforms other baselines as well. Comparing two target habitat models, we find that mapping the result of HR is

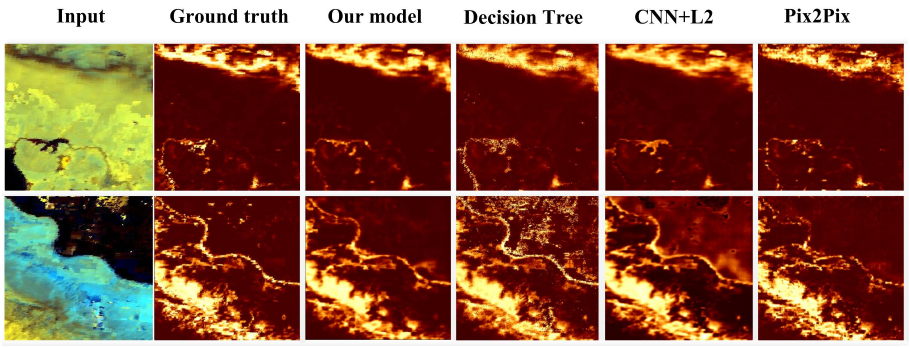

Figure 6: Results on mapping Habitat Suitability Index (HSI) with different methods. The ground truth is produced by the Mahalanobis HSI model with remote sensing images and GPS data. All methods directly map the habitat suitability maps from same remote sensing images. Original probability density maps are colorized with the heat colormap.

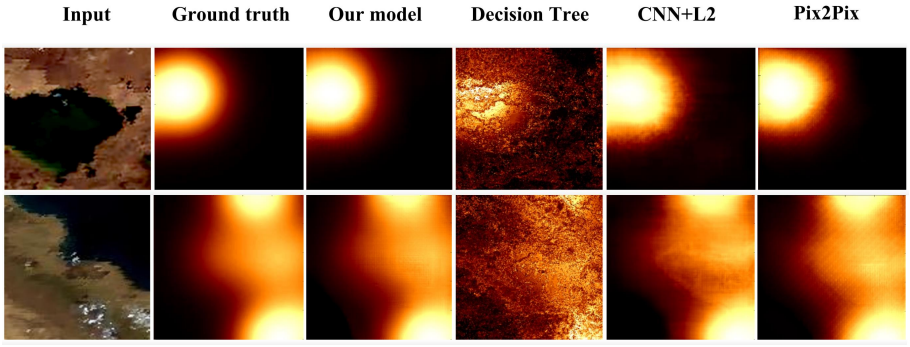

Figure 7: Results on mapping Home Range (HR) with different methods. The ground truth is produced by the kernel HR estimator using only GPS data. We represent input multi-spectral data as true-color images and colorize output probability density maps with the heat colormap.

more susceptible to the lacking of spatial and temporal information.

Discussion

The experimental results indicate that the proposed model is promising in simulating two types of habitat models. For HSI, our model accurately maps the suitability index from source environmental bands, instead of finding their relationship from animal GPS data in advance. Regarding the full-location-based HR, our model succeeds in constructing both the shape and probability of home ranges.

Traditional pixel-based habitat mapping scheme assumes each pixel as an independent vector in multi-dimensional environmental space. It is reasonable when scientists want to identify and explain the habitat characteristic at the pixel level. However, remote sensing imagery is highly structured, and their pixels exhibit strong dependencies. When mapping more complex targets, the neglect of structural information shows its deficiency. As for CGAN-based Image-to-Image translation models, the adversarial loss can be viewed as a high-level goal to train the encoder-decoder structure (Isola et al. 2016). Therefore, they bring better results than ℓ_2 loss. Our experiments confirm the contribution of the adversarial loss in mapping high-quality results.

Producing time-series data from generative adversarial networks is still an open problem. Previous studies have successfully generated delicate video (Saito, Matsumoto, and Saito 2017) and music clips (Dong et al. 2018) by adding a temporal generator. The generated temporal information is used by the following main generator to produce data in specific types. Differently, the temporal patterns in remote sensing imagery are cyclic and simpler than video and music. So we want to directly extract temporal vectors from data and feed them into the generator. Our solution is compatible with most existing adversarial frameworks. It is an efficient implementation for time-series image-to-image translation.

As for remote sensing data, we choose MODIS because of its sufficient reflectance bands. More importantly, our mapping target will barely benefit from high-resolution imagery. Considering the probability map is computed from each pixel of remote sensing images, using high-resolution imagery will bring more computational pressure other than ecological meaning.

Table 1: Mapping accuracy for different methods, evaluated on simulating the Habitat Suitability Index (HSI) model.

Model	Overall			Spatial			Temporal		
	RMSE	MAE	$\mathcal{L}^{SSIM}$	RMSE	MAE	$\mathcal{L}^{SSIM}$	RMSE	MAE	$\mathcal{L}^{SSIM}$
Our model	19.096	11.433	0.473	19.631	11.974	0.481	19.203	11.544	0.462
Decision Tree	26.885	16.712	0.781	26.842	16.917	0.794	27.497	17.335	0.767
CNN + ℓ_2 loss	19.974	14.817	0.685	20.831	15.103	0.691	20.462	15.861	0.674
pix2pix	24.143	15.079	0.627	23.976	14.831	0.625	25.113	15.695	0.619

Table 2: Mapping accuracy for different methods, evaluated on simulating the Home Range (HR) model.

Model	Overall			Spatial			Temporal		
	RMSE	MAE	$\mathcal{L}^{SSIM}$	RMSE	MAE	$\mathcal{L}^{SSIM}$	RMSE	MAE	$\mathcal{L}^{SSIM}$
Our model	15.941	11.126	0.324	17.038	13.973	0.346	16.227	11.098	0.331
Decision Tree	30.673	21.466	0.951	30.784	21.472	0.942	31.016	20.874	0.935
CNN + ℓ_2 loss	18.385	19.837	0.503	21.636	20.738	0.541	19.471	20.613	0.522
pix2pix	17.758	15.447	0.425	19.213	17.932	0.447	19.112	16.147	0.431

Our work still has limitations. One is that the mapping accuracy is highly dependent on the habits of target species. In this example, Bar-headed Geese stay intensively in the lakeshore wetlands and estuaries during the breeding season. The related land covers are mainly shrubland and bare land (Zheng et al. 2018a), which is suitable for mapping the habitat from remote sensing data. Otherwise, the current selection of environmental maps heavily relies on expert knowledge. When mapping the environment-independent models such as HR, different selections will directly influence the mapping accuracy. More improvements are required before applying this approach to rigorous and high-demand analysis.

Conclusion

This paper shows a novel way to simulate probabilistic habitat models with remote sensing imagery and a specially designed time-series image-to-image translation model. Our solution enables scientists to enlarge their wildlife habitat analysis even the radio-tracking data are not enough for large-range and long-term research. The proposed model also provides a new sight to bring temporal information into general image-to-image translation model. So far, few artificial intelligence and deep learning applications have emerged in traditional scientific fields such as animal ecology. These technologies offer more alternatives to solve pending problems. We hope our cross-disciplinary study could inspire more valuable applications and benefit more research communities.

Acknowledgments

We thank Shiming Xiang for many helpful comments. We appreciate the computing resources provided by HTCondor Team in UW-Madison. Research is supported by China Scholarship Council (No.201604910598), Natural Science Foundation of China (61361126011, 90912006), the National R&D Infrastructure and Facility Development Program of China (DKA2018-12-02-XX), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19060205).

References

Bayliss, J. L.; Simonite, V.; and Thompson, S. 2005. The use of probabilistic habitat suitability models for biodiversity action planning. *Agriculture, ecosystems & environment* 108(3):228–250.

Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; and Shah, R. 1994. Signature verification using a" siamese" time delay neural network. In *Advances in neural information processing systems*, 737–744.

Calenge, C. 2006. The package "adehabitat" for the r software: a tool for the analysis of space and habitat use by animals. *Ecological modelling* 197(3):516–519.

Calenge, C. 2011a. Exploratory analysis of the habitat selection by the wildlife in r: the adehabitaths package.

Calenge, C. 2011b. Home range estimation in r: the adehabitathr package. *Office national de la classe et de la faune sauvage: Saint Benoist, Auffargis, France*.

Cappelle, J.; Girard, O.; Fofana, B.; Gaidet, N.; and Gilbert, M. 2010. Ecological modeling of the spatial distribution of wild waterbirds to identify the main areas where avian influenza viruses are circulating in the inner niger delta, mali. *EcoHealth* 7(3):283–293.

Chegoonian, A.; Mokhtarzade, M.; and Valadan Zoej, M. 2017. A comprehensive evaluation of classification algorithms for coral reef habitat mapping: challenges related to quantity, quality, and impurity of training samples. *International Journal of Remote Sensing* 38(14):4224–4243.

Chen, H.; Smith, G.; Zhang, S.; Qin, K.; Wang, J.; Li, K.; Webster, R.; Peiris, J.; and Guan, Y. 2005. Avian flu: H5n1 virus outbreak in migratory waterfowl. *Nature* 436(7048):191–192.

Chopra, S.; Hadsell, R.; and LeCun, Y. 2005. Learning a similarity metric discriminatively, with application to face verification. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, 539–546. IEEE.

Clark, J. D.; Dunn, J. E.; and Smith, K. G. 1993. A multivariate model of female black bear habitat use for a geographic information system. *The Journal of wildlife management* 519–526.

Courty, N.; Flamary, R.; and Ducoffe, M. 2017. Learning wasserstein embeddings. *arXiv preprint arXiv:1710.07457*.

Dong, Z.; Wang, Z.; Liu, D.; Li, L.; Ren, C.; Tang, X.; Jia, M.; and Liu, C. 2013. Assessment of habitat suitability for waterbirds in the west songnen plain, china, using remote sensing and gis. *Ecological engineering* 55:94–100.

Dong, H.-W.; Hsiao, W.-Y.; Yang, L.-C.; and Yang, Y.-H. 2018. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In *Proc. AAAI Conf. Artificial Intelligence*.

Horne, J. S.; Garton, E. O.; Krone, S. M.; and Lewis, J. S. 2007. Analyzing animal movements using brownian bridges. *Ecology* 88(9):2354–2363.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2016. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*.

Justice, C. O.; Vermote, E.; Townshend, J. R.; Defries, R.; Roy, D. P.; Hall, D. K.; Salomonson, V. V.; Privette, J. L.; Riggs, G.; Strahler, A.; et al. 1998. The moderate resolution imaging spectroradiometer (modis): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing* 36(4):1228–1249.

Kobler, A., and Adamic, M. 2000. Identifying brown bear habitat by a combined gis and machine learning method. *Ecological Modelling* 135(2-3):291–300.

Lee, S.; Lee, S.; Song, W.; and Lee, M.-J. 2017. Habitat potential mapping of marten (martes flavigula) and leopard cat (prionailurus bengalensis) in south korea using artificial neural network machine learning. *Applied Sciences* 7(9):912.

Mao, X.; Li, Q.; Xie, H.; Lau, R. Y.; Wang, Z.; and Smolley, S. P. 2016. Least squares generative adversarial networks. *arXiv preprint ArXiv:1611.04076*.

Mirza, M., and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Mohr, C. O. 1947. Table of equivalent populations of north american small mammals. *The American Midland Naturalist* 37(1):223–249.

Nagendra, H.; Lucas, R.; Honrado, J. P.; Jongman, R. H.; Tarantino, C.; Adamo, M.; and Mairota, P. 2013. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators* 33:45–59.

Pastick, N. J.; Jorgenson, M. T.; Wylie, B. K.; Nield, S. J.; Johnson, K. D.; and Finley, A. O. 2015. Distribution of near-surface permafrost in alaska: Estimates of present and future conditions. *Remote Sensing of Environment* 168:301–315.

Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; and Efros, A. A. 2016. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2536–2544.

Rokach, L., and Maimon, O. Z. 2008. *Data mining with decision trees: theory and applications*, volume 69. World scientific.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241. Springer.

Saito, M.; Matsumoto, E.; and Saito, S. 2017. Temporal generative adversarial nets with singular value clipping. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, 5.

Sangkloy, P.; Lu, J.; Fang, C.; Yu, F.; and Hays, J. 2017. Scribbler: Controlling deep image synthesis with sketch and color. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2.

Takekawa, J. Y.; Newman, S. H.; Xiao, X.; Prosser, D. J.; Spragens, K. A.; Palm, E. C.; Yan, B.; Li, T.; Lei, F.; Zhao, D.; et al. 2010. Migration of waterfowl in the east asian flyway and spatial relationship to hpai h5n1 outbreaks. *Avian diseases* 54(s1):466–476.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13(4):600–612.

Worton, B. J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70(1):164–168.

Zhang, W.; Li, X.; Yu, L.; and Si, Y. 2018. Multi-scale habitat selection by two declining east asian waterfowl species at their core spring stopover area. *Ecological Indicators* 87:127–135.

Zheng, R.; Smith, L.; Prosser, D.; Takekawa, J.; Newman, S.; Sullivan, J.; Luo, Z.; and Yan, B. 2018a. Investigating home range, movement pattern, and habitat selection of bar-headed geese during breeding season at qinghai lake, china. *Animals* 8(10):182.

Zheng, R.; Wu, G.; Yan, C.; Zhang, R.; Luo, Z.; and Yan, B. 2018b. Exploration in mapping kernel-based home range models from remote sensing imagery with conditional adversarial networks. *Remote Sensing* 10(11):1722.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017a. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*.

Zhu, J.-Y.; Zhang, R.; Pathak, D.; Darrell, T.; Efros, A. A.; Wang, O.; and Shechtman, E. 2017b. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, 465–476.