Remote Sensing in Ecology and Conservation
Open Access
ZSL LET'S WORK FOR WILDLIFE

ORIGINAL RESEARCH

A convolutional neural network architecture designed for the automated survey of seabird colonies

Hieu Le[1], Dimitris Samaras[1] & Heather J. Lynch[2,3]

[1]Department of Computer Science, Stony Brook University, 263 New Computer Science Building, Stony Brook New York, 11794,
[2]Department of Ecology and Evolution, Stony Brook University, 650 Life Sciences Building, Stony Brook New York, 11794,
[3]Institute for Advanced Computational Science, Stony Brook University, 163 IACS Building, Stony Brook New York, 11794,

Keywords
Adélie penguin, convolutional neural network, high-resolution satellite imagery, prior information

Correspondence
Heather J. Lynch, Department of Ecology and Evolution, Stony Brook University, 650 Life Sciences Building, Stony Brook, NY 11794. Tel: 631 632-9508; Fax: 631 632 7626; E-mail: heather.lynch@stonybrook.edu

Editor: Nathalie Pettorelli
Associate Editor: Tobias Kuemmerle

Received: 19 April 2021; Revised: 29 July 2021; Accepted: 14 September 2021

doi: 10.1002/rse2.240

Abstract

Satellite imagery is now well established as a method of finding and estimating the abundance of Antarctic penguin colonies. However, the delineation and classification of penguin colonies in sub-meter satellite imagery has required the use of expert observers and is highly labor intensive, precluding regular censuses at the pan-Antarctic scale. Here we present the first automated pipeline for the segmentation and classification of seabird colonies in high-resolution satellite imagery. Our method leverages site-fidelity by using images from previous years to improve classification performance but is robust to georegistration artifacts imposed by misalignment between sensors or terrain correction. We use a segmentation network with an additional branch that extracts the useful information from the prior mask of the input image. This prior branch provides the main model information on the location and size of guano in a prior annotation yet automatically learns to compensate for potential misalignment between the prior mask and the input image being classified. Our approach outperforms the previous approach by 44%, improving the average Intersection-over-Union segmentation score from 0.34 to 0.50. While penguin guano remains a challenging target for segmentation due to its indistinct and highly variable appearance, the inclusion of prior information represents a key step toward automated image annotation for population monitoring. Moreover, this method can be adapted for other ecological applications where the dynamics of landscape change are slow relative to the repeat frequency of available imagery and prior information may be available to aid with image annotation.

Introduction

Earth observation (EO) imagery provides new opportunities for monitoring the state of the planet, particularly in the polar regions where logistical challenges preclude direct access to many locations (e.g., LaRue & Knight, 2014; Lynch et al., 2012). The use of satellite imagery to identify penguin colonies in Antarctica extends back to the work of Schwaller et al. (1984) but has seen more active development in the last decade as the availability of satellite imagery continues to grow alongside an interest in monitoring Antarctic species impacted by climate change. Much of this work has focused on the use of medium-resolution Landsat imagery (Fretwell & Trathan, 2009; Lynch & Schwaller, 2014; Schwaller et al., 2013). Algorithms to identify guano in Landsat imagery have been used to detect the presence of penguin colonies (e.g., Borowicz et al., 2018; Schwaller et al., 2013) and estimate colony abundance (Lynch & Schwaller, 2014). High-resolution (sub-meter) commercial satellite imagery has also been used to map penguin colonies (e.g., Fretwell et al., 2012; LaRue et al., 2014; Lynch & LaRue, 2014; Lynch et al., 2012), but the automation of that process has proven much more difficult.

Although penguin guano has a distinct spectral signature that aids in separating it from the surrounding terrain in satellite imagery, penguin colonies can range widely in shape and size (from ~1 to 10^6 m^2; Lynch & LaRue, 2014). At very high resolution, each pixel is so small that the spectral properties of the guano are easily

confused with the spectral properties of common geological (e.g., eroded ridge tops) and biological (e.g., snow algae) features. Guano can be highly variable across sites and even between different images of the same site, as can the background substrate on which the guano is deposited; this heterogeneity makes it difficult to develop a general classification algorithm that can be used reliably across the entire Antarctic (Witharana & Lynch, 2016). Limited training data also make it difficult to overcome scene-level properties, such as atmospheric effects (e.g., clouds, haze) that can change the spectral signature of guano. Notably, human annotators are not strongly influenced by guano color and can easily adjust for clouds, in large part because the guano stain is relatively stable in location and shape between years and experts can use that prior expectation to quickly discard other non-essential information. Our aim in this work was to leverage the implicit use of prior expectation in the manual classification of imagery for penguin colonies to improve classification performance.

Automating imagery classification

Because of the pre-existing volume of imagery to be processed, and the rate at which new Antarctic imagery is captured, existing classification pipelines are not yet robust enough for full automation, requiring time-consuming review by experts familiar with each penguin colony. As a result, we must consider methods that involve super-pixel features and spatial context, both of which are critical for manual annotation of sub-meter commercial imagery. Imagery classification is an active field (e.g., Arefin et al., 2020; Baghbaderani et al., 2020; Garnot et al., 2020; Lary et al., 2018; Xie et al., 2017; Zheng et al., 2020; Zhu et al., 2017), particularly with the integration of new methods, such as 'deep learning', from computer vision and machine learning. As a result, deep learning approaches such as convolutional neural networks (CNNs) are rapidly taking their place as invaluable tools alongside more traditional pixel-based classification methods (Zhu et al., 2017). CNNs are computational models composed of multiple convolutional layers (Lecun et al., 1989). Each layer of a CNN consists of multiple learnable filters that allow the CNN to capture important spatial information from the input signals. In the last decade, CNNs have dramatically improved many computer vision benchmark records, especially thanks to the implementation of CNNs on GPUs (Ciresan et al., 2012). While CNNs often require a large number of data points to train, they can be particularly valuable when feature characteristics appear at the super-pixel scale or rely on contextual clues elsewhere in the scene. Some target classes, however, remain challenging even using state-of-the-art CNN architectures; features may be difficult to classify accurately using CNNs if they are rare and training sets are thus unavoidably small, or if they are similar to other objects in the scene. Animals are frequently both rare and camouflaged, and even aggregations of animals may be small relative to the resolution of the sensor or the area to be searched.

Despite these challenges, CNNs have been developed to monitor wildlife populations from satellite imagery, including the identification of pack-ice seals (Goncalves et al., 2020) and whales (Borowicz et al., 2019) in Antarctica. Although the areal extent of a penguin colony is much larger than that of an individual seal or whale, their identification has proven surprisingly challenging and an automated pipeline for extracting colony area has lagged these other applications. Fortunately, the size, shape and extent of penguin colonies evolve slowly over time due to their long lifespan and high nest site fidelity. As a result, each image is similar to the previous image and prior classification can be used to greatly improve classification accuracy. Here we focus on the use of prior information to improve the classification performance of a CNN-based approach to identifying penguin guano in sub-meter resolution satellite imagery. The overall shape and location of the penguin colony presented in the prior mask can directly aid the segmentation model. In fact, image segmentation with priors (Gulshan et al., 2010; Isack et al., 2018; Le et al., 2016, 2017; Luo et al., 2019) has been shown to effectively improve object segmentation accuracy and precision. In the context of remote sensing, there have been several efforts to incorporate a shape prior into a segmentation model. For example, a contour model was used in Han and Wu (2017) for river image segmentation and Maggiori et al. (2015) used a segmentation model with a shape prior (Gorelick et al., 2014) for semantic segmentation of satellite imagery. However, these methods cannot be applied to cases, like the identification of penguin guano, where the target shape and size varies significantly and where misregistration may corrupt the alignment between each image and its prior mask.

Using prior information to improve segmentation

In previous work (Le et al., 2019), we trained an Adélie penguin colony segmentation model using a semi-weakly supervised framework that used the medium-resolution Landsat-based masks from Lynch and LaRue (2014) as a weak annotation to train a classifier for penguin guano in high-resolution (Worldview-2, Worldview-3, QuickBird-2, Geoeye) imagery. Specifically, we used the guano classifications from Lynch and LaRue (2014) to identify 2044

unlabeled high-resolution images over Adélie penguin colonies and to serve as weak annotation for the location and approximate extent of guano in each image. We combined these weakly labeled images with a small number (n = 18) of manually annotated high-resolution images to train our model. Because penguins are highly site faithful, the size, shape and location of penguin colonies does not change dramatically from one year to the next. Accordingly, the annotation of penguin guano in a previous year's image provides important information on areas more likely to contain penguin guano. However, the use of prior information becomes challenging because natural fluctuations in the appearance of the colony are convolved with georegistration errors or orthorectification artifacts that result in mismatches between the weak-annotation masks and the input images. Thus, in Le et al. (2019), we approximated the area of the penguin colony in the training image from the weakly annotated masks and enforced a constraint that the segmentation network should learn how to generate output masks with the same number of pixels as the corresponding masks created through weak annotation. In this way, the use of weakly annotated masks provided a useful training signal allowing the network to learn useful features from a large number of unlabeled images and only a minimal set of labeled images. While largely successful, this approach did not attempt to use the prior information directly to incorporate the shape and the location of the penguin colony into the segmentation model.

Here we describe a custom-designed CNN architecture that takes full advantage of a prior mask by injecting the prior mask into the segmentation model via a prior branch. This branch extracts the global geometric structure and location of the penguin colony from the prior mask. While these masks provide highly salient information about locations where penguin guano is more likely to be found, as well as information on colony shape, the actual location of guano in the image to be classified may be significantly different either due to (1) natural changes in the size and shape of the penguin colony over time, or (2) georegistration errors or orthorectification artifacts. While guano evolution may be slow from one year to the next, the prior mask may be several years (and up to a decade) from the image to be classified, compounding the potential mismatch due to the dynamics of the colony. Orthorectification artifacts arise primarily when the terrain model used for orthorectification changes over time, which can create differences in guano shape and size between two images processed using different terrain models. Some examples of prior masks are shown in Figure 1. On the one hand, these prior masks hint at the overall shape of the colonies and their approximate location. On the other hand, they could introduce misleading information to the segmentation model since only some parts of the masks overlap the ground-truth masks. Thus, we seek a model that is flexible enough to extract useful information from the prior masks but does not cause the model to ignore strong guano signatures in the actual input image. By training this branch together with the main segmentation branch in an end-to-end segmentation framework, we introduce flexibility in the use of the prior mask to compensate for the potential misalignment between the prior mask and the input image. This design overcomes image registration mismatches, which are common to many applications, and improves classification accuracy across a range of image conditions.

As of August 2018, we had 19 404 high-resolution images over Adélie penguin colonies available to be annotated, which at 1 h/scene (conservative) would require over 9 years of full-time annotation by a penguin expert. In a similar use case involving pack-ice seals (Goncalves et al., 2020), we demonstrated that automated algorithms for target detection in high-resolution imagery using only a single GPU can reduce processing time by >95% when compared to an expert annotator. Thus, our automated approach not only speeds up the process of monitoring penguins by satellite imagery but is in fact absolutely *necessary* for the kind of routine cost-effective monitoring required to understand penguin dynamics across the entire Antarctic. While our particular use case is focused on a specific application, small training datasets, image heterogeneity and poor discrimination between target and non-target features are widespread challenges. Fortunately, however, many EO applications (such as ours) benefit from dynamics that are slow relative to the repeat frequency of the available imagery. The use of prior knowledge naturally extends to the classification of an imagery time series, which in the aggregate can be used to understand the underlying dynamics of landscape change. Additional applications of this method would include studies of changing forest composition (Hansen et al., 2013), forest cover (Eidenshink et al., 2007), fires and related disturbances (Huang et al., 2010; Liu et al., 2006), landslides (Tralli et al., 2005) and urban sprawl (Grădinaru et al., 2019; Yang & Lo, 2002).

Materials and Methods

Data

In this work, we use multi-band high-resolution satellite imagery (Worldview-2 and -3, Quickbird-2 and -3 and Geoeye) provided by Maxar, Inc. (Westminster, Colorado, USA), for training and testing our model. All multi-band images were converted to RGB images by selecting the respective bands and normalizing the intensity values to

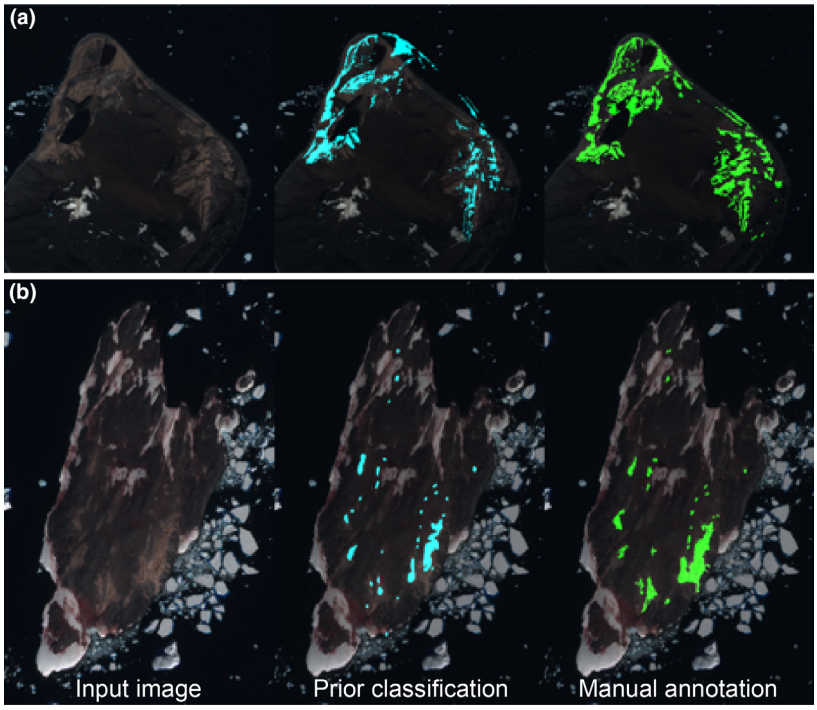

Figure 1. Prior masks and hand-annotated masks for two penguin colonies as seen in Worldview-2 imagery (top row (A): Paulet Island, bottom row (B): Chappel Island) in our testing set. From left to right are as follows: input images, the prior masks overlaid on top of the input images and the hand-annotated masks of the penguin colonies. The prior masks provide information on the overall shape and approximate location of the penguin colonies. Satellite imagery copyright Maxar, Inc. 2021.

(0–255). We hand-annotated a set of 90 images, which are split into 18 training images from 18 different penguin colonies and 72 testing images from 21 different penguin colonies (Fig. 2; image details in Data S1). Colony size ranged from ~1 m^2 to over 10 000 m^2. There were 32 penguin colonies in total. In the testing set, there were seven colonies that had images included in the training set and the remaining 25 were new locations not represented in the training set. All images were collected during the austral summer when penguins are nesting and guano is visible.

Processing

To prepare imagery for use in our model, we split each image and its annotation mask into patches of size 386 × 386 with 50% overlap between neighboring patches. Each training patch is randomly cropped into the size of 256 × 256, followed by several standardized image augmentation techniques, that is, randomly flipped, rotated by a random (−10, 10) degree, and random color jittering. For testing, we split each image into patches of size 256 × 256 with 50% overlap between neighboring patches.

At the core of our image processing pipeline is our segmentation model PenguinNet (Fig. 3). PenguinNet is designed based on a U-Net (Ronneberger et al., 2015), which is composed of a down-sampling part and an up-sampling part. The input is first processed by a sequence of down-sampling blocks that could extract features across different scales, including both global contexts and local cues that are useful to detect penguin guano. These features are gradually up-scaled and processed to predict an output heatmap via a sequence of up-sampling blocks. The architecture also sends the output of down-sampling blocks directly to corresponding up-sampling blocks (i.e. skip connections) to be combined with up-sampled outputs that facilitate the learning of useful patterns such as the identity function.

To incorporate the information from the prior mask, we custom designed an additional down-sampling prior branch with similar architecture as the main down-sampling branch but with fewer convolutional filters. The prior branch takes as input the prior mask only; features extracted from this branch serve as additional cues for the network to detect the penguin guano, which are merged into the main branch. The position at which we merge prior features with the main network is a

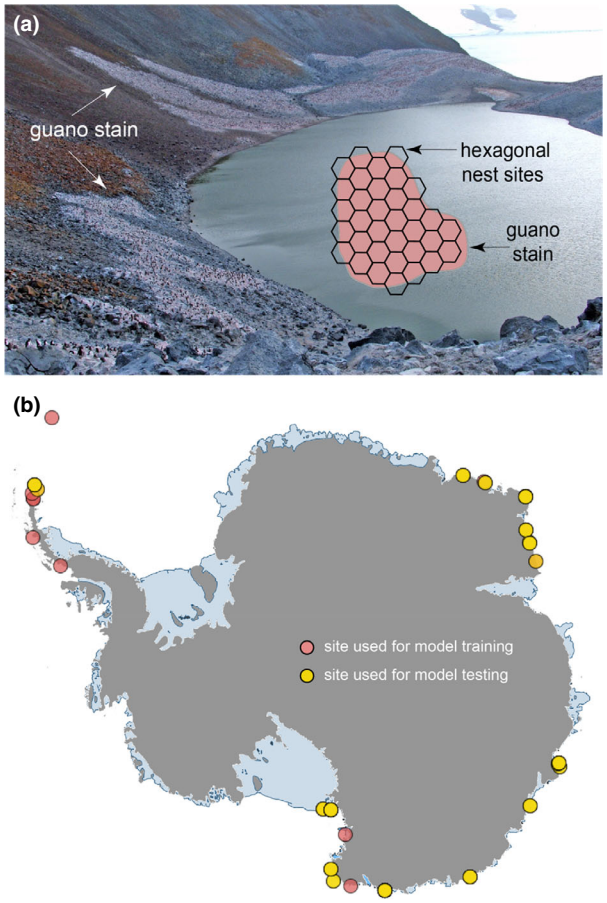

Figure 2. (A) Photograph of an Adélie penguin colony on Paulet Island showing the pinkish-white guano stain used to identify penguin colonies from satellite imagery. Inset: Diagram illustrating the approximately hexagonal packing of penguins within the guano stain as illustrated by the pink polygon. (B) The locations of penguin colonies around Antarctica used for training and testing of the guano detection algorithm.

hyper-parameter of our model, determining the number of down-sampling operations and, consequently, the optimal level of detail for features from prior images. This branch is colored as blue with the length n = 3 in Figure 3. While merging prior features later can ameliorate the difficulties imposed by image-mask misalignment, down-sampling the prior features too much causes a loss of information on the location of the colony and degrades segmentation performance. To obtain the output segmentation mask, we first crop the input image into patches of size 256 × 256 with 50% overlap between neighboring patches. We use the trained network to obtain a prediction mask for each patch and average all overlapped patch predictions at each pixel to obtain the prediction mask for each image.

Similar to our previous method (Le et al., 2019), we train our network in a semi-supervised learning manner. We use a small set of 18 images with hand-annotated guano areas and a set of 2044 images with their prior masks to train our network. Among the 2044 automatically amassed images, there are images without visible guano areas due to bad weather conditions such as heavy snow, clouds, shadows, or because the timing of the image was not well aligned with the period of guano visibility. We first trained a patch-level classification CNN using patches extracted from areas with and without visible guano within the 18 hand-annotated images to filter out images without visible guano areas from our training set. Whenever hand-annotated masks are available for a scene, we train the network to output an identical pixel mask. For scenes where only prior masks are available, the network is forced to generate prediction masks such that the sum values of the prediction masks equal the sum of values for the prior masks (i.e. the prediction and ground-truth contain the same area of guano).

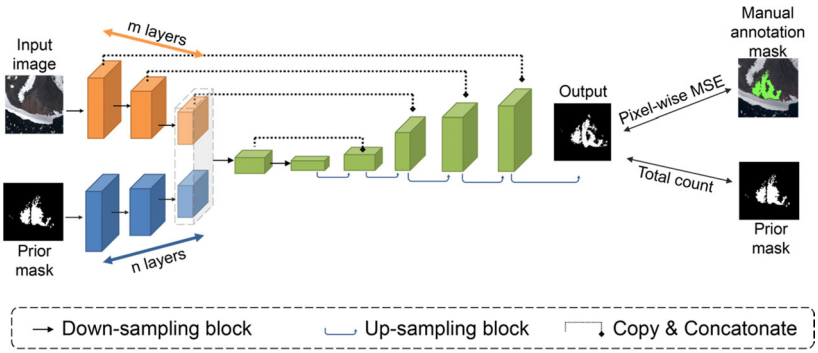

Figure 3. PenguinNet architecture. Our network is based on a U-Net architecture. We input the prior mask into a separate CNN branch constructed by *n* down-sampling blocks. Similarly, the input image is processed separately using a CNN branch constructed by *m* down-sampling blocks. The outputs of these two branches are concatenated, followed by several down-sampling and up-sampling blocks to predict the final output segmentation mask.

Results

We evaluated our method on 72 testing images from 21 different penguin colonies in which guano stains were hand-annotated (Table 1). The new model here described outperforms both a baseline U-Net model (trained using only 18 images from the fully supervised training set) and the 'no prior' model (like the new model, trained with the 18 images from the fully supervised training set and the 2044 images from our database of penguin colony images with misaligned masks) from Le et al. (2019) (Table 1), using mean Intersection-Over-Union (mIoU) and pixel-wise accuracy to evaluate the output segmentation masks. Note that a simple baseline where we use 4-channel input images consisting of three RGB color channels and one channel for the prior mask can be considered as one specific configuration of our model, that is, a model with a 0-layer prior branch that injects the information at the 0-th layer of the main branch. The results for our proposed model in this paper are computed from a model with a prior branch consisting of 5 downsampling blocks and merges into the main branch at the 5th layer.

Because each image has slightly different orthorectification, and the priors do not exactly align with the actual guano in the image being classified, the priors provide only basic information on the shape and location of the penguin colonies. Our model capitalizes on the information inherent in these prior images and outperforms the benchmark ('non-prior') model (Fig. 4), which lacks the branch for the prior guano mask, in various scenarios. As seen in Figure 4A, our model correctly identifies the majority of guano areas whose locations are suggested by the prior mask. By contrast, the non-prior model from Le et al. (2019) is only able to identify guano in a localized area. Importantly, our model does not simply copy pixel-level information from the prior mask but rather incorporates global information on shape and size obtained from the prior mask to improve the segmentation. In Figure 4B and C, our model almost perfectly predicts the exact mask that was annotated by an expert while the non-prior model over-estimates the guano area because it is confused by spectrally similar areas that lead, in the absence of any auxiliary information on the guano location, to false positives. In Figure 4D, there are areas in the scene where both the texture and colors are similar to guano-covered areas. The non-prior model misclassifies them as guano while also fails to identify the correct guano areas. By contrast, the prediction from our prior-based model largely overlaps with the human annotation. Figure 4E shows a challenging case where the guano areas are only barely visible due to poor weather conditions. These kinds of images would generally not be considered viable for abundance estimation and were not included in the training set. While the non-prior model fails to detect any guano areas, the new model performs surprisingly well in this case, suggesting that the prior-based model generalizes to a wider set of conditions than its non-prior counterpart.

Table 1. Segmentation performance (mIoU and pixel-wise accuracies for guano and non-guano classes) of different models on the testing set of penguin colony images.

Model	mIoU	Guano accuracy	Non-guano accuracy
Baseline U-Net	0.26	29.7%	99.7%
No prior (Le et al., 2019)	0.36	50.9%	98.5%
Stacking prior masks on top of input images (no separated CNN branch)	0.42	66.8%	98.7%
Using a separated CNN branch for prior masks (proposed)	0.50	71.1%	98.7%

The testing set includes 72 images from 21 penguin colonies. mIoU, mean Intersection-over-Union; CNN, convolutional neural network.

As a second mechanism to evaluate model performance, we manually graded segmentation performance on 162 images from six penguin colonies with well-known colony configurations. This method allows us to quickly assess the overall performance of the segmentation method without time-consuming guano annotation. We rated each output guano mask with a score ranging from 1 to 5: 1 being 'poor' and 5 being 'perfect' (Figure 5). A rating of 4 indicated segmentation en par with manual human annotation, with only very small errors similar to those common when manually annotating images. Table 2 summarizes the performances of our model in comparison with the baseline model that does not use a prior branch. As can be seen, our model outperforms the baseline model on all sites by at least 0.5 points on this five-point scale. Several images were unusable for reasons related to cloud cover, recent snow, or a severe mismatch between the timing of the image and the appearance of guano at the colony. Unsurprisingly, the percentage of unusable images peaks at the beginning (September; 44%) and end (March, 67%) of the austral summer breeding season, whereas only 5% of December images were considered unusable for classification. Conditional on being a usable image, however, there was no strong trend in segmentation performance from December through March; all 4 months yielded strong (average score >4.0) segmentation performance.

Overall, our model outperformed the non-prior model by a large margin in terms of both mIoU and visual inspection. It effectively extracts the crucial information

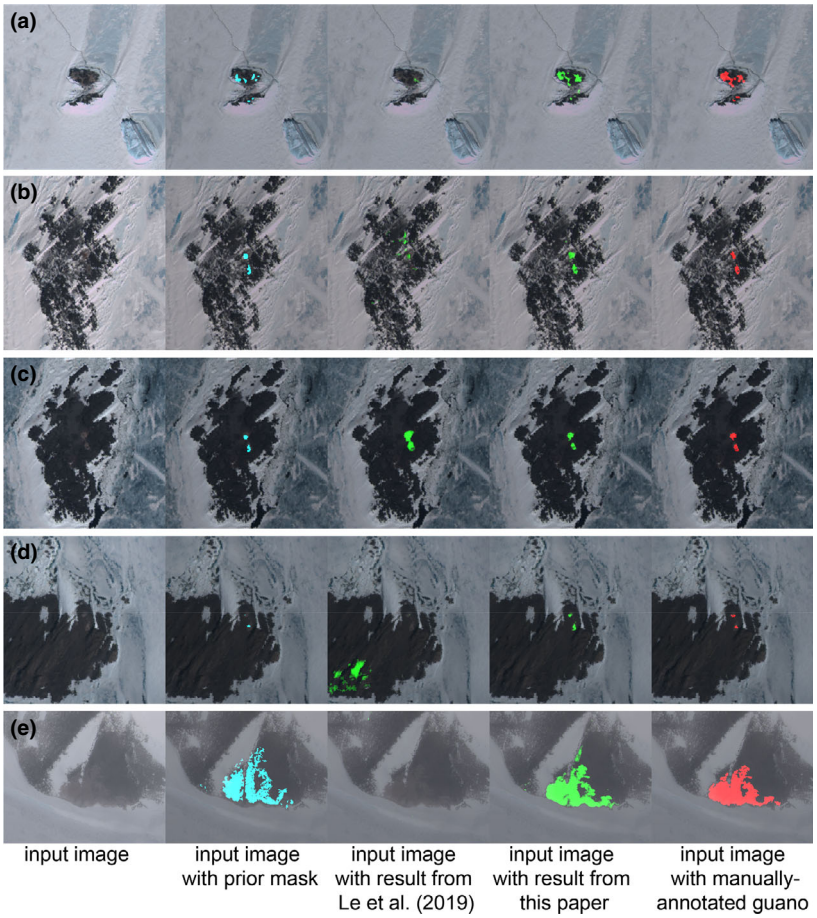

Figure 4. Segmentation results of our proposed model in comparison with our previous method (Le et al., 2019) that does not include a branch for the prior mask for five images of penguin colonies: (A) Worldview-2 image of Chick Islands on 2 December 2015, (B) Worldview-2 image of Balaena Islands on 23 November 2015, (C) Worldview-2 image of Balaena Islands on 19 January 2016, (D) Worldview-2 image of McDonald Point on 16 January 2016 and (E) Worldview-3 image of Arthurson Ridge on 24 November 2015. The prior mask is shown in blue, the segmentation results are shown in green (both that from Le et al. [2019] and the newer model being presented here) and the manually annotated mask in red. Satellite imagery copyright Maxar, Inc. 2021.

from the prior masks to better localize and segment the guano areas. As can be seen in Figures 4 and 5, our model works well even in the cases where the prior masks significantly misalign with the actual guano, showing that the model does not simply just copy the information from the prior masks but rather learns how to incorporate the information contained therein with other visual cues from the input image. This is best illustrated in Figure 6 where we visualize segmentation outputs for different images of the colony at Arthurson Ridge (ARTH). It should be noted that the outputs of the model change according to the changes of the penguin colony while using the same prior mask but the approach can introduce failures, particularly when the prior mask is badly misaligned (Figure 7).

Discussion

We report on a novel approach to improve classification for highly challenging segmentation applications in remote sensing. Our application of this method to penguin colonies incorporates challenges common to many remote sensing applications. The spectral signature of a penguin colony is faint, highly variable, and has indistinct edges. Guano color depends on penguin diet and can range from deep pink to white, and the background substrate is highly variable. Even repeated images at a single location can look very different depending on weather events like snow or differences in the timing of the image relative to the species' breeding phenology. In addition, high-resolution satellite imagery is limited and manual

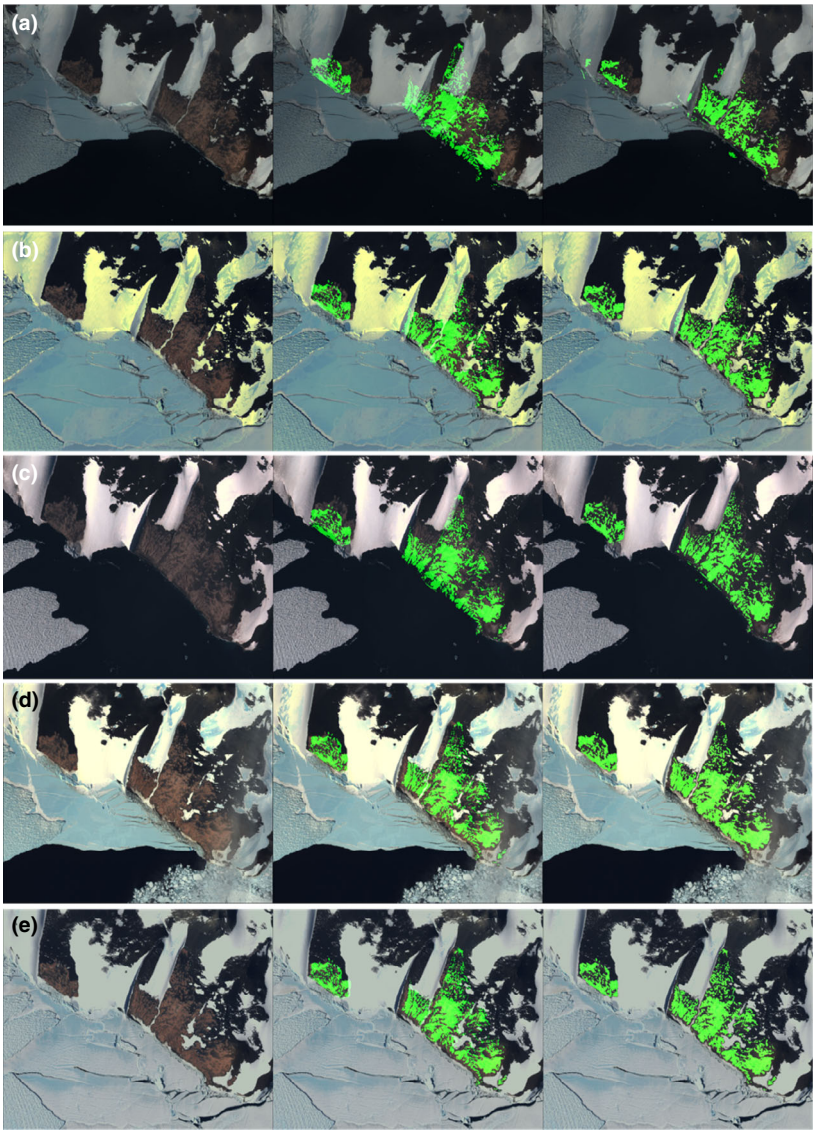

Figure 5. Segmentation outputs for different images of the colony at Cape Crozier. Each row represents a different input image from the lowest to the highest classification score (A–E representing 1–5, respectively): (A) Quickbird-2 image from 16 December 2009, (B) Worldview-2 image from 30 November 2011, (C) Worldview-2 image from 31 January 2011, (D) Quickbird-2 image from 28 November 2009 and (E) Worldview-2 image from 12 December 2011. Within each row, the leftmost image is the input image, the middle image and rightmost images show the input image overlaid with the prior mask and the final segmentation, respectively. Satellite imagery copyright Maxar, Inc. 2021.

guano annotation exceptionally time consuming, which limits the size of training data available for CNN-based classification strategies. Another challenge is the near impossibility of obtaining ground validation coincident with a satellite image, since it is not possible to precisely control the timing of tasking requests and cloud cover is a frequent obstacle. For this reason, our goal here was to develop a system that comes as close as possible to matching the judgment of an expert human annotator. While these challenges present real technical barriers to an automated classification approach, the classification of penguin guano has one significant advantage, which is that the guano stain is largely stable in size, shape and location from one season to the next. For this reason, information from a prior classified image can be highly informative and can improve segmentation and classification performance.

One challenge with incorporating prior classification is that orthorectification can create significant misalignments between the prior masks and the actual locations

Table 2. Average scores of the outputs of our model in comparison with the baseline model that does not use the prior branch.

Site	Number of images	Model without prior (Le et al., 2019)	Model with prior (proposed)
ARTH	80	2.4	3.1
BEAG	6	2.5	3.3
CROZ	47	3.5	4.0
HERO	5	1.8	4.4
HOPE	9	3.2	4.1
BIRD	15	2.5	4.0
All	162	2.8	3.6

The methods are tested on the images of six penguin colony sites: Arthurson Ridge [ARTH], Beagle Island [BEAG], Cape Crozier [CROZ], Heroina Island [HERO], Hope Bay [HOPE] and Cape Bird [BIRD]. Each output guano mask is graded from 1 to 5 with 1 being 'poor' and 5 being 'excellent'.

of guano in the images. For this reason, simply stacking prior masks on the input images does not yield a significant improvement. However, a separate prior branch can learn to extract the information from the prior mask that could most benefit the classification. This design is flexible and can deal with these misalignment issues effectively in our penguin use case. There are multiple ways that our framework can be extended. For example, an interactive framework with a human in the loop would allow for test-time modification on the prior masks. We also can use image registration to reduce the misalignment between the prior mask and the testing image. Improved correction of atmospheric effects and/or color normalization methods would further improve the model's performance and generalizability to new scenes.

While the model performance likely still falls short of what would be required for large-scale automated monitoring of penguin colonies, our inclusion of a prior mask greatly improved the classification success when compared to a previously published model and represents a general strategy to improve classification for the classification of features than evolve slowly relative to the repeat interval of the imagery available. While a limited training dataset is, in the near term at least, unavoidable, we anticipate that injecting additional sources of auxiliary information may further improve classification performance. For example, penguins have specific terrain requirements for nesting (e.g., McDowall & Lynch, 2017) and the recent availability of digital elevation models for the Antarctic (Howat et al., 2019) presents the opportunity to incorporate terrain as an auxiliary layer in the classification CNN. Offering a significant improvement over Le et al. (2019), our model represents a new benchmark for the automated delineation of Adélie penguin colonies in Antarctica and provides a new approach for segmentation of images where the target image is largely stable over time.

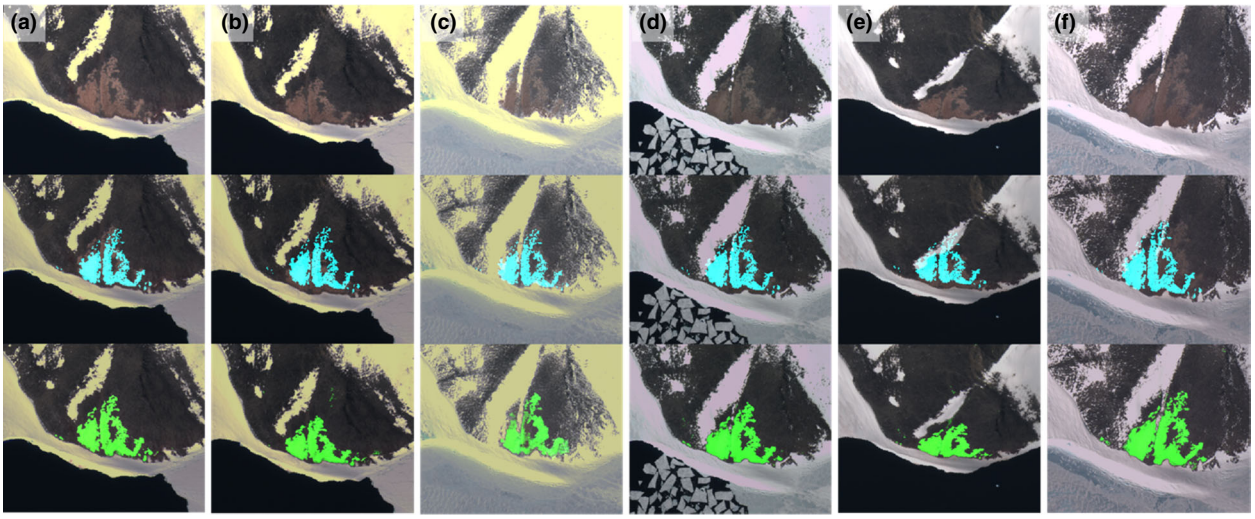

Figure 6. Examples from Arthurson Ridge in which misalignments between the target image and the image from which the prior mask was extracted can lead to poor segmentation results. Each row represents a different input image: (A) Worldview-2 image from 15 February 2015, (B) a second but different Worldview-2 image from 15 February 2015, (C) Worldview-2 image from 23 October 2015, (D) Worldview-2 image from 3 December 2015, (E) Worldview-2 image from 4 February 2016 and (F) Worldview-3 image from 20 November 2014. Within each column, the topmost image is the input image, and the middle image and bottom images show the input image overlaid with the prior mask and the final segmentation, respectively. Satellite imagery copyright Maxar, Inc. 2021.

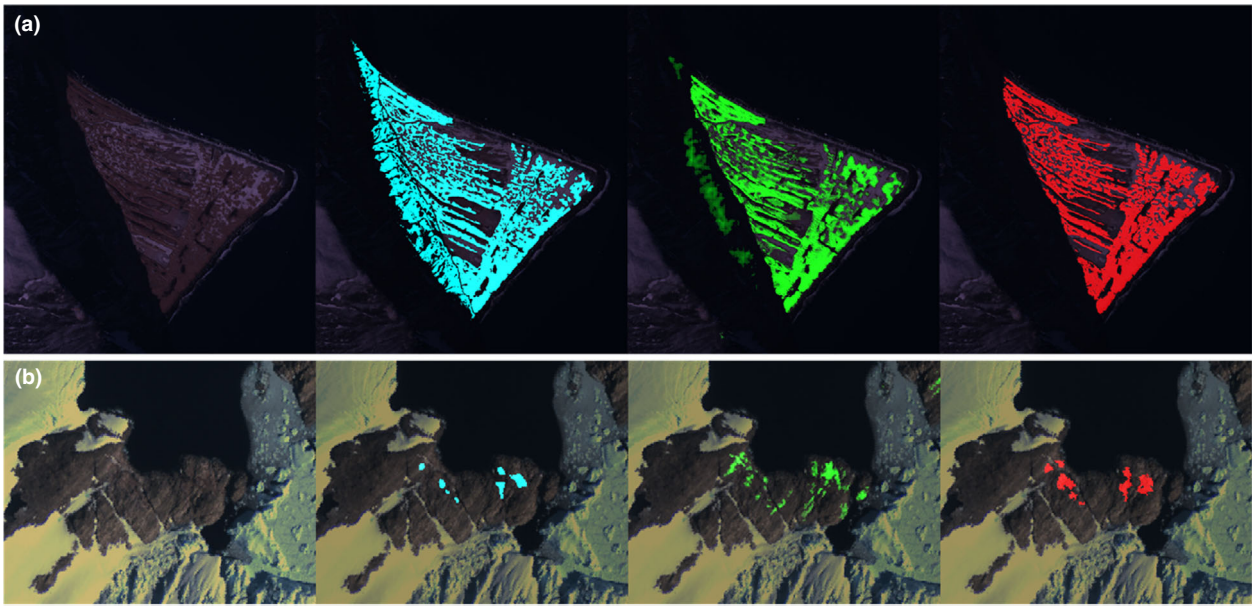

Figure 7. Two examples in which the segmentation failed: (A) Worldview-2 image on 16 February 2011 at Cape Adare and (B) Worldview-2 image on 9 February 2016 at Forbes Glacier. The input image is in the first column, the prior mask is shown in blue (second column), the segmentation results from the current model are shown in green (third column) and the manually annotated mask in red (fourth column). Satellite imagery copyright Maxar, Inc. 2021.

Acknowledgments

We thank the Institute for Advanced Computational Science, National Geographic/Microsoft's AI for Earth program and the National Science Foundation EarthCube program (Award 1740595) for funding this work. Geospatial support for this work was provided by the Polar Geospatial Center under NSF-OPP awards 1043681 and 1559691.

REFERENCES

Arefin, M.R., Michalski, V., St-Charles, P., Kalaitzis, A., Kim, S., Kahou, S. et al. (2020) Multi-image super-resolution for remote sensing using deep recurrent networks. In: Boult, T., Medioni, G. & Zabih, R. (Eds.) *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Los Alamitos, CA: IEEE Computer Society, pp. 816–825.

Baghbaderani, R.K., Qu, Y., Qi, H. & Stutts, C. (2020) Representative-discriminative learning for open-set land cover classification of satellite imagery. 2020 European Conference on Computer Vision (ECCV).

Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V. et al. (2019) Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLoS One*, **14**, e0212532.

Borowicz, A., McDowall, P., Youngflesh, C., Sayre-McCord, T., Clucas, G., Herman, R. et al. (2018) Multi-model survey of Adélie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. *Scientific Reports*, **8**, 3926.

Ciresan, D., Meier, U. & Schmidhuber, J. (2012) Multi-column deep neural networks for image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z.L., Quayle, B. & Howard, S. (2007) A project for monitoring trends in burn severity. *Fire Ecology*, **3**, 3–21.

Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B., Ratcliffe, N. et al. (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One*, **7**, e33751.

Fretwell, P.T. & Trathan, P.N. (2009) Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography*, **18**, 543–552.

Garnot, V.S., Landrieu, L., Giordano, S. & Chehata, N. (2020) Satellite image time series classification with pixel-set encoders and temporal self-attention. In: Boult, T., Medioni, G. & Zabih, R. (Eds.) *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA: IEEE Computer Society, pp. 12322–12331.

Goncalves, B., Spitzbart, B. & Lynch, H.J. (2020) SealNet: a fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, **239**, 111617.

Gorelick, L., Veksler, O., Boykov, Y. & Nieuwenhuis, C. (2014) Convexity shape prior for segmentation. ECCV.

Grădinaru, S.R., Kienast, F. & Psomas, A. (2019) Using multi-seasonal Landsat imagery for rapid identification of abandoned land in areas affected by urban sprawl. *Ecological Indicators*, **96**, 79–86.

Gulshan, V., Rother, C., Criminisi, A., Blake, A. & Zisserman, A. (2010) Geodesic star convexity for interactive image segmentation. In: Darrell, T., Hogg, D. & Jacobs, D. (Eds.) *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society, pp. 3129–3136.

Han, B. & Wu, Y. (2017) A novel active contour model based on modified symmetric cross entropy for remote sensing river image segmentation. *Pattern Recognition*, **67**, 396–409.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.S.S., Tyukavina, A. et al. (2013) High-resolution global maps of 21st-century forest cover change. *Science*, **342**, 850–853.

Howat, I.M., Porter, C., Smith, B.E., Noh, M.-J. & Morin, P. (2019) The reference elevation model of Antarctica. *The Cryosphere*, **13**, 665–674.

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z. & Vogelmann, J.E. (2010) An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, **114**, 183–198.

Isack, H.N., Gorelick, L., Ng, K., Veksler, O. & Boykov, Y. (2018) K-convexity shape priors for segmentation. ECCV.

LaRue, M.A. & Knight, J. (2014) Applications of very high resolution imagery in the study and conservation of predators in the Southern Ocean. *Conservation Biology*, **28**, 1731–1735.

LaRue, M.A., Lynch, H.J., Lyver, P., Barton, K., Ainley, D.G., Pollard, A.M. et al. (2014) Establishing a method to estimate Adélie penguin populations using remotely-sensed imagery. *Polar Biology*, **37**, 507–517.

Lary, D.J., Zewdie, G.K., Liu, X., Wu, D., Levetin, E., Allee, R.J. et al. (2018) Machine learning applications for earth observation. In: Mathieu, P.P. & Aubrecht, C. (Eds.) *Earth observation open science and innovation*. New York: Springer International Publishing, pp. 165–218.

Le, H., Gonçalves, B., Samaras, D. & Lynch, H. (2019) Weakly labeling the Antarctic: the penguin colony case. In: Davis, L., Torr, P. & Zhu, S.C. (Eds.) *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Los Alamitos, CA: IEEE Computer Society.

Le, H., Nguyen, V., Yu, C. & Samaras, D. (2016) Geodesic distance histogram feature for video segmentation. ACCV.

Le, H., Yu, C., Zelinsky, G.J. & Samaras, D. (2017) Co-localization with category-consistent features and geodesic distance propagation. In: Chellappa, R., Zhang, Z. & Hoogs, A. (Eds.) *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Los Alamitos, CA: IEEE Computer Society, pp. 1103–1112.

LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E. et al. (1989) Backpropagation applied to hand-written zip code recognition. *Neural Computation*, **1**, 541–551.

Liu, D., Kelly, M. & Gong, P. (2006) A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote Sensing of Environment*, **101**, 167–180.

Luo, S., Tai, X., Huo, L., Wang, Y. & Glowinski, R. (2019) Convex shape prior for multi-object segmentation using a single level set function. In: Lee, K.M., Forsyth, D., Pollefeys, M. & Tang, X. (Eds.) *ICCV 2019*. Los Alamitos, CA: IEEE Computer Society.

Lynch, H.J. & LaRue, M.A. (2014) First global survey of Adélie penguin populations. *The Auk*, **131**, 457–466.

Lynch, H.J. & Schwaller, M.R. (2014) Mapping the abundance and distribution of Adélie penguins using Landsat-7: first steps towards an integrated multi-sensor pipeline for tracking populations at the continental scale. *PLoS One*, **9**, e113301.

Lynch, H.J., White, R., Black, A.D. & Naveen, R. (2012) Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, **35**, 963–968.

Maggiori, E., Tarabalka, Y. & Charpiat, G. (2015) Optimizing partition trees for multi-object segmentation with shape prior. BMVC.

McDowall, P. & Lynch, H.J. (2017) Ultra-fine scale spatially-integrated mapping of habitat and occupancy using structure-from-motion. *PLoS One*, **12**, e0166773.

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M. & Frangi, A. (Eds.) *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Cham: Springer, pp. 234–241.

Schwaller, M.R., Benninghoff, W.S. & Olson, C.E. Jr. (1984) Prospects for satellite remote sensing of Adélie penguin rookeries. *International Journal of Remote Sensing*, **5**, 849–853.

Schwaller, M.R., Southwell, C.J. & Emmerson, L.M. (2013) Continental-scale mapping of Adélie penguin colonies from Landsat imagery. *Remote Sensing of Environment*, **139**, 353–364.

Tralli, D.M., Blom, R.G., Zlotnicki, V., Donnellan, A. & Evans, D.L. (2005) Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *Photogrammetry & Remote Sensing*, **59**, 185–198.

Witharana, C. & Lynch, H.J. (2016) An object-based image analysis approach for detecting penguin guano in very high spatial resolution satellite images. *Remote Sensing*, **8**, 375.

Xie, D., Zhang, L. & Bai, L. (2017) Deep learning in visual computing and signal processing. *Applied Computational Intelligence and Soft Computing*, **2017**, 1320780.

Yang, X. & Lo, C.P. (2002) Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing*, **23**, 1775–1798.

Zheng, Z., Zhong, Y., Wang, J. & Ma, A. (2020) Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4095–4104.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F. et al. (2017) Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, **5**, 8–36.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Detailed information on satellite images used for model testing and training and model performance on each image.

