Remote Sensing in Ecology and Conservation
Open Access

ORIGINAL RESEARCH

Deep learning detects invasive plant species across complex landscapes using Worldview-2 and Planetscope satellite imagery

Thomas A. Lake, Ryan D. Briscoe Runquist & David A. Moeller

Department of Plant and Microbial Biology, University of Minnesota, 1479 Gortner Ave, St. Paul Minnesota, 55108, USA

Keywords
Convolutional neural network, *Euphorbia esula*, invasive species, phenology, recurrent neural networks, satellite image time series

Correspondence
Thomas A. Lake and David A. Moeller, Department of Plant and Microbial Biology, University of Minnesota, 1479 Gortner Ave, St. Paul, MN 55108. Tel: 612-624-1037; E-mail: lakex055@umn.edu; moeller@umn.edu

Editor: Temuulen Sankey
Associate Editor: Yinghai Ke

Received: 24 November 2021; Revised: 11 May 2022; Accepted: 20 May 2022

doi: 10.1002/rse2.288

Remote Sensing in Ecology and Conservation 2022;**8** (6):875–889

Abstract

Effective management of invasive species requires rapid detection and dynamic monitoring. Remote sensing offers an efficient alternative to field surveys for invasive plants; however, distinguishing individual plant species can be challenging especially over geographic scales. Satellite imagery is the most practical source of data for developing predictive models over landscapes, but spatial resolution and spectral information can be limiting. We used two types of satellite imagery to detect the invasive plant, leafy spurge (*Euphorbia virgata*), across a heterogeneous landscape in Minnesota, USA. We developed convolutional neural networks (CNNs) with imagery from Worldview-2 and Planetscope satellites. Worldview-2 imagery has high spatial and spectral resolution, but images are not routinely taken in space or time. By contrast, Planetscope imagery has lower spatial and spectral resolution, but images are taken daily across Earth. The former had 96.1% accuracy in detecting leafy spurge, whereas the latter had 89.9% accuracy. Second, we modified the CNN for Planetscope with a long short-term memory (LSTM) layer that leverages information on phenology from a time series of images. The detection accuracy of the Planetscope LSTM model was 96.3%, on par with the high resolution, Worldview-2 model. Across models, most false-positive errors occurred near true populations, indicating that these errors are not consequential for management. We identified that early and mid-season phenological periods in the Planetscope time series were key to predicting leafy spurge. Additionally, green, red-edge and near-infrared spectral bands were important for differentiating leafy spurge from other vegetation. These findings suggest that deep learning models can accurately identify individual species over complex landscapes even with satellite imagery of modest spatial and spectral resolution if a temporal series of images is incorporated. Our results will help inform future management efforts using remote sensing to identify invasive plants, especially across large-scale, remote and data-sparse areas.

Introduction

Invasive plant species are reshaping landscapes globally, harming agricultural and ecological systems and generating substantial economic losses (Diagne et al., 2021; Marbuah et al., 2014; Pimentel et al., 2005). Knowledge of the distribution and abundance of invasive plants is critical for reducing the costs of invasion (Bradley et al., 2019; Sofaer et al., 2018). To obtain such information, scientists and land managers traditionally rely on in situ measurements of population size, density or growth rates via field surveys, which can be prohibitively expensive and require considerable effort (Gaskin et al., 2020). However, the rate of invasive species proliferation across the globe has made in situ measurements largely untenable (Seebens et al., 2017). Thus, alternative techniques are urgently needed to capture information on invasive plant species more quickly and efficiently.

Remote sensing is perhaps the most promising alternative means to survey plant species (Bradley, 2014; Huang

& Asner, 2009; Kattenborn et al., 2021; Pettorelli et al., 2014). However, species-level detection is a challenging task for remote sensing because there may not be sufficient signal to discriminate among individual species (Isaacson et al., 2012). Additionally, variation in population density, growth form and phenology across landscapes further complicates detection (Kattenborn et al., 2019). Modern satellite technology (e.g. Worldview-2) can capture sub-meter resolution images at a single time period. Concurrently, Earth observation missions (e.g. Planetscope) can capture large swaths of the Earth's surface continuously at a rapid return rate with slightly lower image resolution (Curnick et al., 2021; Roy et al., 2021). These types of satellite images have potential use for detection and monitoring of individual plant species, however, few studies to date have shown high accuracy in single-species detection across complex landscapes and at broad geographic scales.

Plant species with distinctive spectral, textural and phenological characteristics make excellent targets for remote sensing (Bradley, 2014; Clinton et al., 2010; Huang & Asner, 2009). Differences in chlorophyll content and leaf chemistry affects the reflectance, transmission and absorption of light, which can be used to distinguish plant species (Asner & Martin, 2008; Clark et al., 2005; Martin et al., 1998). Likewise, leaf pigmentation and the timing of leaf senescence has long been used as a phenological signal to discriminate among canopy tree species (Dymond et al., 2002; Isaacson et al., 2012; Wolter et al., 1995). Traditional statistical classifiers (e.g. support vector machines, random forests) incorporate spectral and phenological information with a set of manually selected indices (Irons & Petersen, 1981; Xue & Su, 2017) (termed 'feature engineering'). However, selecting a set of the most appropriate features can be challenging, as it requires knowledge of a species' biology and interactions with the electromagnetic signal measured by a sensor (Kattenborn et al., 2021). By contrast, the recent development of deep learning neural networks largely circumvents the feature engineering process and often outperforms shallower machine learning algorithms for vegetation classification (Kattenborn et al., 2021). The neural network can itself learn which features are best suited to predict a plant species of interest (Kattenborn et al., 2021).

Deep learning algorithms have demonstrated state-of-the-art performance in classification tasks across disciplines (Krizhevsky et al., 2017; Reichstein et al., 2019). The key advantages of these algorithms are their ability to learn complex patterns in data (i.e. nonlinear relationships), synthesize large volumes of data and circumvent the feature engineering process (Goodfellow et al., 2016; LeCun et al., 2015). Convolutional neural networks (CNNs) are among the most common deep learning algorithms applied in remote sensing and are increasingly used for vegetation classification (Kattenborn et al., 2021; Ma et al., 2019; Sefrin et al., 2020), but have only recently been used to detect plant species (Kattenborn et al., 2020, 2021; Rist et al., 2019; Schiefer et al., 2020). CNNs learn from the spectral, spatial and contextual information within an image yet have rarely incorporated temporal information. Long Short-Term Memory (LSTM) networks are a type of deep learning algorithm developed to analyze temporal sequences of data (Hochreiter & Schmidhuber, 1997; Yu et al., 2019). In the context of plant communities, LSTMs can use phenological signal to discriminate among coexisting species that may be indistinguishable at single time points. While LSTMs are only beginning to be evaluated for detecting plant species (Larson & Tuor, 2021), recent findings suggest that incorporating phenological trends improves crop classification (Ghosh et al., 2021; Mazzia et al., 2020; Rußwurm & Körner, 2018a) and invasive species detection (Dahal et al., 2022; Weisberg et al., 2021). Overall, CNNs and LSTMs may be particularly promising for species identification in complex landscapes given the increasing availability of satellite image time series.

Deep learning algorithms use millions of parameters making model interpretation challenging (Samek et al., 2019), and few remote sensing studies interpret deep learning models beyond basic summary statistics (Campos-Taberner et al., 2020). However, researchers are now starting to investigate the inner workings of these models by manipulating model inputs (i.e. images) (Kattenborn et al., 2021; Rußwurm & Körner, 2018a). For example, Mäyrä et al. (2021) systematically occluded portions of satellite imagery to identify which regions of an image were influential for tree species classification. By exploring similar techniques, researchers can better understand why deep learning algorithms make predictions and determine which features are essential for plant species detection.

Our study focused on the detection of the invasive plant species leafy spurge (*Euphorbia virgata*; Euphorbiaceae) across urban, natural and rural landscapes using deep learning models and satellite imagery. Our first objective was to ask how the spatial and spectral resolution of satellite images affected the performance of CNNs at detecting of leafy spurge. We constructed and evaluated two CNN models; the first using Worldview-2 images (resolution 0.5 m; eight spectral bands) taken during the peak flowering period and the second using Planetscope images (resolution 3 m; four spectral bands) captured during the same time period. Our second objective was to ask whether the addition of a time series of Planetscope satellite images with an LSTM could capture phenological trends and improve the accuracy of detecting leafy spurge. Last, we probed the inner-workings of our models to determine where and how the models succeeded and failed across the study area. We applied a series of data manipulations to Worldview-2 and Planetscope satellite imagery by modifying spectral bands and phenological periods to determine how these features were crucial for leafy spurge prediction.

Materials and Methods

Study system

Leafy spurge (*Euphorbia virgata*; Euphorbiaceae; Levin & Gillespie, 2017, also referred to as *E. esula*) was introduced to North America from Eurasia in the early 1800s, likely as a seed contaminant (DiTomaso, 2000; Duncan et al., 2004; Wallace et al., 1992). Multiple introductions of leafy spurge by the late 1800s allowed for widespread invasion throughout North America (Dunn, 1979). Since introduction, leafy spurge has spread rapidly with infested acreage roughly doubling every decade through the early 2000s (Leistritz et al., 1992; Leitch et al., 1994). Leafy spurge now occupies nearly 2 million hectares in the Northern Great Plains of North America and causes more than $130 million in reduced grazing and rangeland activity annually (Bangsund et al., 1999; DiTomaso, 2000; Duncan et al., 2004; Leitch et al., 1994). Several features make leafy spurge suitable for remote sensing: it is widespread and abundant, tends to grow in monocultures and produces yellow-green flowering bracts in early spring that have a distinct spectral and phenological signature driven by lower chlorophyll a, b and greater carotenoids relative to co-occurring native vegetation (Fig. S1) (Hooge Hom et al., 2020; Hunt et al., 2004).

Leafy spurge is an herbaceous, C3 perennial forb that spreads via ballistically dispersed seeds and vegetative reproduction (Fig. 1). Leafy spurge grows in diverse habitats: soils vary from sandy to heavy clay and it has a wide climatic distribution (Lym, 2005). Established plants are among the first to emerge from dormancy in the spring, which confers a competitive advantage over native broadleaf plants and grasses (Bangsund et al., 1999; Belcher & Wilson, 1989; Leistritz et al., 1992).

Study workflow

We first undertook extensive population surveys to establish a ground truth dataset. Second, we obtained and preprocessed Worldview-2 and Planetscope satellite imagery for deep learning analyses. Third, we constructed, trained and evaluated three deep learning models based on the Worldview-2 and Planetscope satellite data. In addition, we examined the inner-workings of our models by visualizing several internal model layers, which allowed us to determine the features (e.g. regions, patterns, textures) that were used in model inference. Last, we manipulated spectral and phenological information in satellite imagery to interrogate their importance for leafy spurge prediction (workflow outlined in Fig. 2).

Population censuses

In 2019 and 2020, we documented leafy spurge throughout six areas of interest (AOIs) in Minnesota, USA, which is situated within the north-central portion of the species' range (Fig. 1). We selected six AOIs totaling 161 km^2 that had a high frequency of populations and comprised a representative sample of regional land use and habitat contexts. Within each AOI, we conducted an extensive set of ground surveys by travelling on-foot and by vehicle. Where populations were found, we recorded presence and boundaries using a handheld smartphone GPS and a Garmin 64ST GPS (Garmin Ltd) by placing waypoints at individual occurrence sites. In total, we catalogued 1,860 georeferenced points within populations, which were used to digitize leafy spurge on a land cover map.

Worldview-2 satellite images and preprocessing

We obtained Worldview-2 images (Maxar Technologies) captured in early June, 2020 for each of our six AOIs (Table S2), coinciding with the peak flowering period of leafy spurge in our region. The Worldview-2 satellite contains a panchromatic band (0.5-m resolution) and eight spectral bands: coastal, blue, green, yellow, red, red-edge, near infrared 1 and near infrared 2 at 1.8-m resolution. We applied radiometric, atmospheric, geometric and pansharpening corrections prior to deep learning analyses (Method S1).

Planetscope satellite images and preprocessing

We obtained 12 Planetscope images between 2 May 2020 and 30 July 2020 for each of our six AOIs from Planet Labs (Planet Labs, Inc., 2020). We selected cloud-free images that captured the phenological sequence of early growth and flowering of leafy spurge within each of our AOIs and prioritized imagery from the Planetscope PS2 sensor due to differences in the spectral response between the Planetscope PS2.SD and PSB.SD sensors (Frazier & Hemingway, 2021). We matched the extent of the Planetscope images with the Worldview images to directly compare predictions between image sources. Planetscope images were obtained as orthorectified analytics-ready

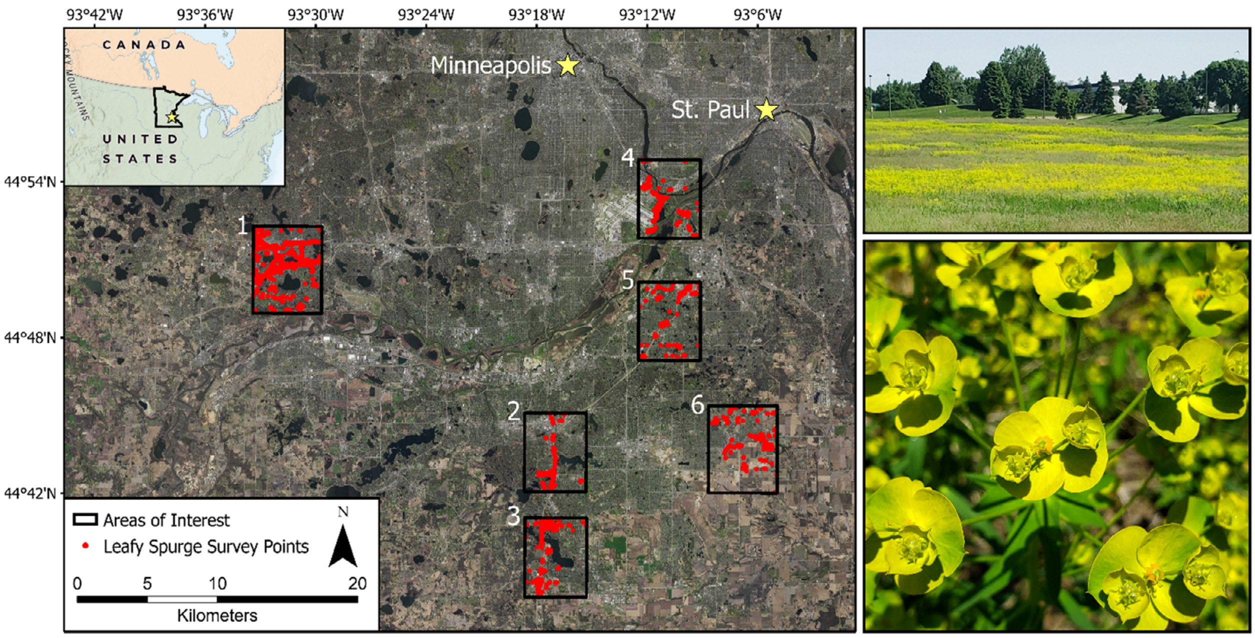

Figure 1. Left panel: Map of six study areas (black boxes) in the Twin Cities metropolitan area, Minnesota, USA that were used to develop deep learning models. Red points indicate georeferenced occurrences of leafy spurge, and stars indicate major city centers. Right panel: Photos of a large leafy spurge population captured during the flowering period, and an individual plant with bright yellow-green flowering bracts.

surface reflectance. Planetscope imagery has four spectral bands including blue, green, red and near infrared with a pixel size of 3-m. Image metadata are available in supplement (Table S3).

Labeling land cover map

We obtained a 1-m resolution land cover map of our study area (Host et al., 2016) to use as a ground truth for training deep learning models (Fig. S2). A land cover map is required for our deep learning approach, termed image segmentation (i.e. pixel-wise image labeling), in which a model learns to associate satellite image pixels with land cover classes. We digitized leafy spurge onto the land cover map following georeferenced coordinates collected during ground surveys. For inaccessible survey areas (e.g. private lands), we digitized leafy spurge with visual interpretation of the Worldview-2 satellite imagery. We reclassified the land cover map into six primary groups of interest: native vegetation (e.g. deciduous/coniferous trees, emergent/forest wetlands, grasses/shrubs), buildings (e.g. structures over 3 m tall), roads, water, agriculture (e.g. annual/perennial crops) and leafy spurge (Table S1). We resampled the land cover map to 0.5 m for the Worldview-2 imagery and 3 m for the Planetscope imagery using bilinear interpolation (ESRI, 2020). In total, we digitized 1,383 distinct leafy spurge populations ranging from 6 to 15 311 m^2 within our study region.

U-net convolutional neural network

Deep learning architectures consist of many individual neurons arranged in a hierarchical structure to learn progressively abstract features from data (Goodfellow et al., 2016). The state-of-the-art deep learning architecture for image segmentation is the convolutional neutral network (CNN), where neurons are arranged in the dimensions of an image (i.e. height, width) (Krizhevsky et al., 2017; LeCun et al., 2015). We built our CNN using a U-net architecture (Ronneberger et al., 2015) to segment leafy spurge and five other land cover classes from satellite imagery. The U-net architecture contains an encoder and decoder stage, whereby the encoder extracts features (e.g. lines, edges, shapes, textures) from an image using a series of convolution, activation and pooling layers, and the decoder projects features back to the original image dimensions using a series of mirrored upsampling, concatenation and convolution steps (Ronneberger et al., 2015).

We used the *segmentation-models* package (https://github.com/qubvel/segmentation_models) to select an untrained U-net architecture (Yakubovskiy, 2019). We selected the SE-ResNet50 U-net architecture because it was among the top performers in the 2017 ImageNet Large Scale Visual Recognition Challenge (Hu et al., 2020). This U-net combines recent innovations from ResNet blocks (He et al., 2015) and Squeeze-and-Excitation blocks (Hu et al., 2020), which have been

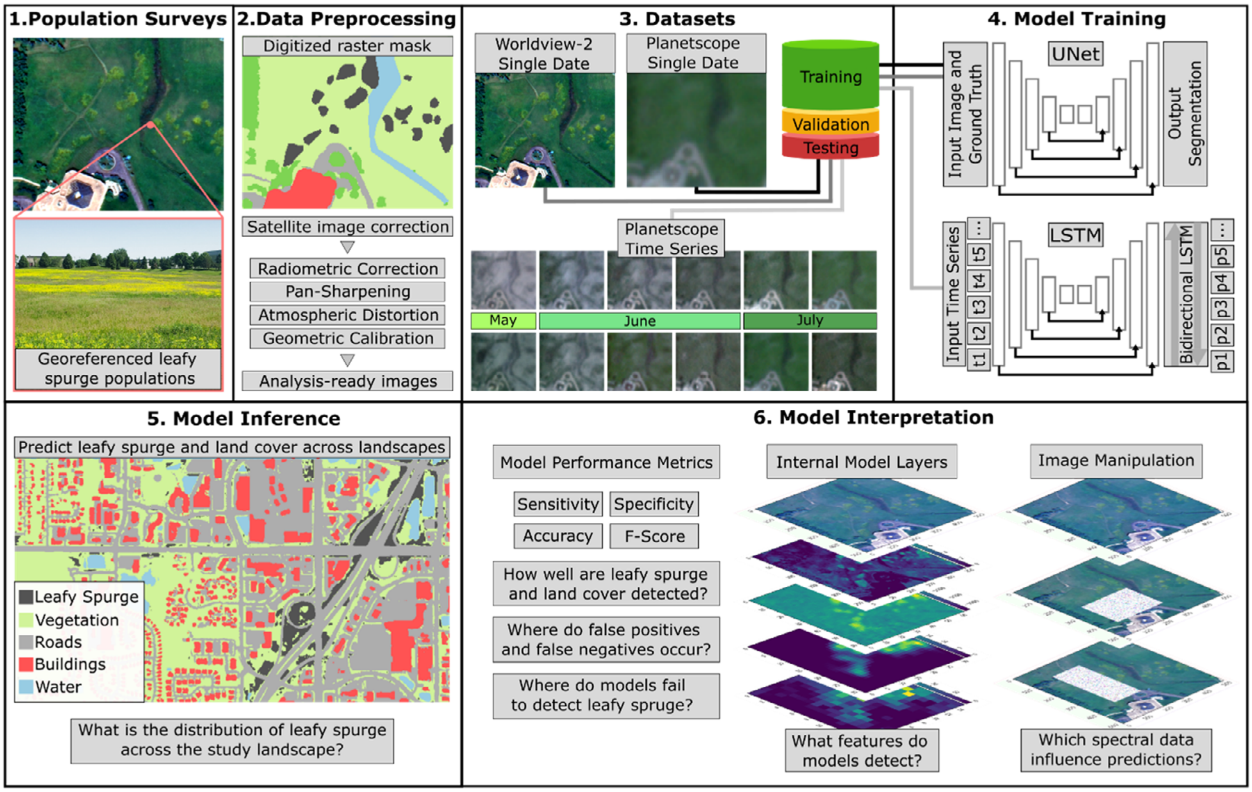

Figure 2. Workflow diagram summarizing six methodological steps taken to construct, train and evaluate deep learning models for the remote sensing of leafy spurge from satellite imagery. (1) Population surveys characterized leafy spurge for deep learning analyses. (2) Data preprocessing involved digitizing surveyed leafy spurge populations onto a land cover raster mask, and correcting several aspects of satellite imagery. (3) Datasets included single-date Worldview-2 and Planetscope satellite imagery, as well as a time series of Planetscope satellite imagery. The single-date imagery sources (black line, dark gray line) were divided into training, validation and testing sets and subsequently used in CNN deep learning models, and the time series imagery source (light gray line) was used for the LSTM model. (4) Model training required iterating through batches of satellite images and land cover masks to learn the association between the input images and target outputs. (5) Model inference included prediction of leafy spurge and land cover classes across the study landscape. (6) Model interpretation steps aimed to evaluate performance and interrogate how and why models detect leafy spurge.

found to boost model performance (Hu et al., 2020). The ResNet block adds residual connections, which allows for information to be passed between layers in the U-net encoder (He et al., 2015). The Squeeze-Excitation block adjusts (i.e. squeezes) features according to the activations of previous layers in the encoder, allowing the network to learn to associate specific features with specific regions of interest (Hu et al., 2020). The final layer of the U-net, here the SoftMax function, normalizes the outputs on the scale of [0,1] to produce a predicted probability for each class per pixel (Fig. S3). We used the above U-net CNN architecture for the Worldview-2 U-net model (WV-CNN) and Planetscope U-net model (PS-CNN).

LSTM model

To learn phenological trends from the time series of satellite images, we modified our U-net to include a Long Short-Term Memory (LSTM) layer following the initial CNN and prior to the SoftMax layer (Sefrin et al., 2020). The LSTM model takes as input a chronological sequence of Planetscope images (i.e. a 5D tensor with dimensions [samples, timesteps, image rows, image columns, image channels]) and outputs a prediction for each class at each timestep (i.e. a 5D tensor with dimensions [samples, timesteps, image rows, image columns, classes]). The LSTM layer learns the phenological sequence in Planetscope images by relying on information from both previous and current timesteps, which influences a prediction at the current timestep. We applied the ConvLSTM2D block to allow the LSTM to use two-dimensional images as input (Shi et al., 2015). The ConvLSTM2D block consists of a convolutional operation and an LSTM operation, which are applied to the chronological sequence of Planetscope images. The convolutional operation is performed to extract spatial features from the input image, while the LSTM operation is performed to extract temporal features. Additionally, we implemented a bidirectional LSTM layer such that the model learns phenology in forward and reverse chronological order, which can improve LSTM performance (Graves et al., 2005; Schuster & Paliwal, 1997) (Fig. S4). We used the above architecture for the Planetscope LSTM (PS-LSTM) model.

Model datasets and training

We developed all models in Python v3.6.9 with Keras v2.2.4 (Chollet, 2015) and TensorFlow v2.0.0 (Abadi et al., 2016). For the Worldview-2 imagery, we generated a total of 2604 512 × 512-pixel image tiles in our study areas. We used 512 × 512 image tiles to include a sufficient contextual area for model training. For the Planetscope imagery, we generated a total of 12 804 128 × 128-pixel image tiles across 12 dates. We randomly partitioned each dataset into 60% for model training, 20% for validation, and withheld 20% for an independent testing set to evaluate model performance. Within the entire dataset, native vegetation represented ~50% of total pixels, whereas leafy spurge comprised ~1% of total pixels, presenting a potential class imbalance problem (Table S1). Following Cabezas et al. (2020), we calculated a class weights vector based on the inverse frequency in which classes occur in the full dataset, allowing the model to upweight correct predictions of leafy spurge relative to native vegetation and other classes. For model training, we used the Adaptive Moment Estimation optimizer (Kingma, 2014), the categorical cross-entropy loss function, and a learning rate of 1e-4. We trained each model for 100 epochs, where each epoch comprised 200 samples per batch with a batch size of eight. We applied random image augmentations, including 0/90/180/270-degree rotations, which are shown to assist in model generalization (Cabezas et al., 2020; Shorten & Khoshgoftaar, 2019). All code is available online (https://github.com/lake-thomas/spurge-remote-sensing).

Model performance metrics

We assessed model performance for each class based on the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). We calculated overall accuracy as the proportion of correctly identified pixels (TP + TN/TP + TN + FP + FN) to identify the probability that an individual class is correctly classified. However, overall accuracy may be prone to bias in cases of imbalanced data. We further calculated class-wise specificity, or the ability to discriminate true negatives as (TN/TN + FP), and sensitivity (also termed recall), or the ability to predict all true positives (TP/TP + FN), to understand performance in relation to false positives and false negatives. In other words, models with high specificity do not overpredict leafy spurge (i.e. correctly identify sites where leafy spurge is known to be absent), and models with high sensitivity demonstrate that they correctly detect most instances of leafy spurge. Additionally, we calculated the F2-score, which is derived from precision (TP/TP + FP) and sensitivity as: [5*precision*sensitivity)/(4* (precision + sensitivity)]. We selected the F2 score to place more importance on sensitivity relative to precision, as performance metrics for invasive species should weigh false negatives greater than false positives, since the potential economic cost of falsely declaring an invasive species absent is high (Schmidt et al., 2012).

Model interpretation

Explaining model predictions is pertinent for understanding a model's limitations, however, this is not straightforward in deep learning models (Samek et al., 2019). One common strategy is to expose the intermediate layers of the network and investigate which features influence predictions (Rußwurm & Körner, 2018b). We visualized intermediate model layers to uncover which features (e.g. regions, patterns, textures) are used in model inference. However, visualizing intermediate layers only reveals which features are detected on a per-layer basis, but not what drives overall predictions.

To gain insight into how spectral data drive overall predictions, we applied a series of data manipulations to the spectral information from satellite imagery. Briefly, for each spectral band of a satellite image, we replaced a test square of an image with spectral data sampled from the empirical distribution for that band derived from either leafy spurge or vegetation. We then changed the mean of the distribution and resampled the data. Last, we determined if the probability of predicting leafy spurge changed based on the mean of the sampled data and which spectral band was being manipulated (Fig. 3). We then compared the peak spectral reflectance values between leafy spurge and vegetation to define the extent to which a spectral band distinguishes each class.

Further, to understand which phenological periods were most important for predicting leafy spurge, we applied a series of data manipulation to the time series imagery in the Planetscope LSTM model. We divided the sequence of 12 Planetscope images into three sections of four images; early, mid and late season. We then manipulated each of the three sections separateley by applying random Gaussian noise to each image to disrupt the phenological signal present in a sequence of satellite images. Substantial changes to leafy spurge predictions after manipulating either early, mid or late season indicate that time period is crucial for accurate predictions.

Results

Worldview-2 U-net model

The U-net model trained using the Worldview-2 imagery (WV-CNN) produced an overall accuracy of 96.1% for the leafy spurge class when evaluated using the withheld testing dataset (Table 1). Additionally, the WV-CNN model had high specificity (96.1%) and sensitivity (89.7%) for the leafy spurge class (Table 1). By contrast, the model had lower precision (11.9%) and F2 scores (38.8%) due to an elevated proportion of false positives (i.e. predicting vegetation pixels as leafy spurge) compared to true positives for the leafy spurge class (Fig. 4; Table 1).

Planetscope U-net model

U-net models developed with Planetscope imagery (PS-CNN) produced an overall accuracy of 89.9% for the leafy spurge class, driven by moderate sensitivity (61.0%) and high specificity (90.1%) (Table 2). Similar to the WV-CNN, excess false positives relative to true positives for the leafy spurge class reduced precision (3.6%) and F2 scores (14.6%) (Table 2; Fig. 4).

Planetscope LSTM model

Using a time series of Planetscope satellite imagery with an LSTM (PS-LSTM) resulted in an overall accuracy of 96.3% for the leafy spurge class. Increased accuracy was achieved via gains in both sensitivity (66.3%) and specificity (96.5%) relative to the PS-CNN model (Table 3). The precision (11%) and F2 score (33.0%) were greater relative to the PS-CNN model due to a reduced proportion of false positives relative to true positives (Table 3; Fig. 4).

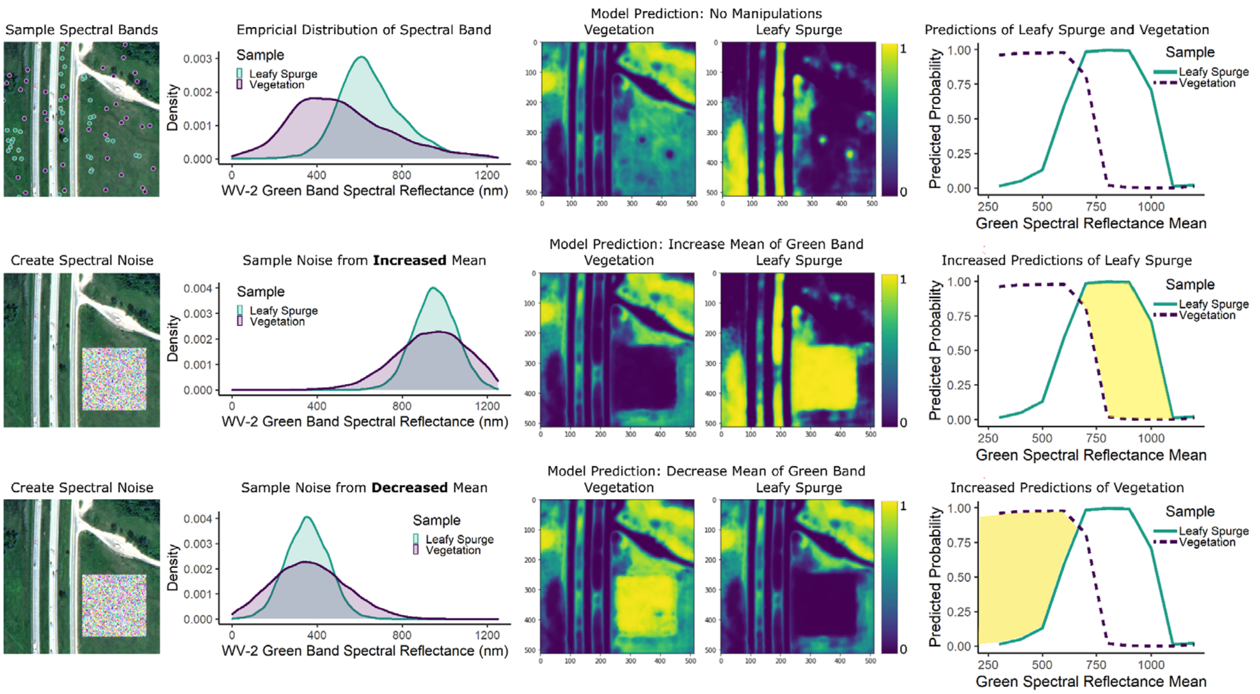

Figure 3. Spectral data manipulations demonstrate that changes in spectral information influence model predictions of leafy spurge and vegetation. Top row: Spectral data were sampled from 10 000 points randomly drawn from the leafy spurge and vegetation classes. Empirical distributions of each spectral band (shown here, the green band from Worldview-2) were created for leafy spurge and vegetation. The mean spectral reflectance of the green band for leafy spurge samples was higher relative to vegetation samples. Model predictions with no spectral manipulations show the predicted probability of leafy spurge and vegetation, with higher probabilities in lighter yellow shades. Middle and bottom rows: Spectral noise was generated by randomly sampling spectral reflectance band values from each of the leafy spurge and vegetation classes. Spectral noise was then applied to a patch on Worldview-2 and Planetscope images, and models were tasked with predicting images with noise. The mean of the spectral noise was increased and decreased across a regular interval to identify whether the model was sensitive to changes in spectral information for correct predictions of vegetation and leafy spurge. Increasing the mean value of the green spectral band leads to a greater predicted probability for leafy spurge, and decreasing the mean value of the green spectral band leads to a greater predicted probability for vegetation.

Table 1. Worldview-2 U-net (WV-CNN) model evaluation metrics demonstrating model performance.

WV-CNN	Accuracy	Sensitivity	Specificity	Precision	F2 Score
Vegetation	0.878	0.827	0.949	0.957	0.850
Buildings	0.958	0.908	0.962	0.637	0.837
Roads	0.936	0.843	0.958	0.827	0.840
Water	0.983	0.970	0.984	0.831	0.938
Agriculture	0.990	0.937	0.995	0.936	0.937
Leafy Spurge	0.961	0.897	0.961	0.119	0.388

Models for land cover classification

Our models performed well when classifying buildings, roads, water and agriculture (Tables 1–3; Fig. 4). Building segmentation using the WV-CNN model outperformed either PS-CNN and PS-LSTM models across all metrics, likely because Worldview-2 imagery had the spatial resolution to identify small or partially hidden buildings. The PS-LSTM model improved the classification accuracy of classes with distinct temporal changes (Table 3). Agriculture classification for the PS-LSTM model relative to the PS-CNN model improved slightly in overall accuracy

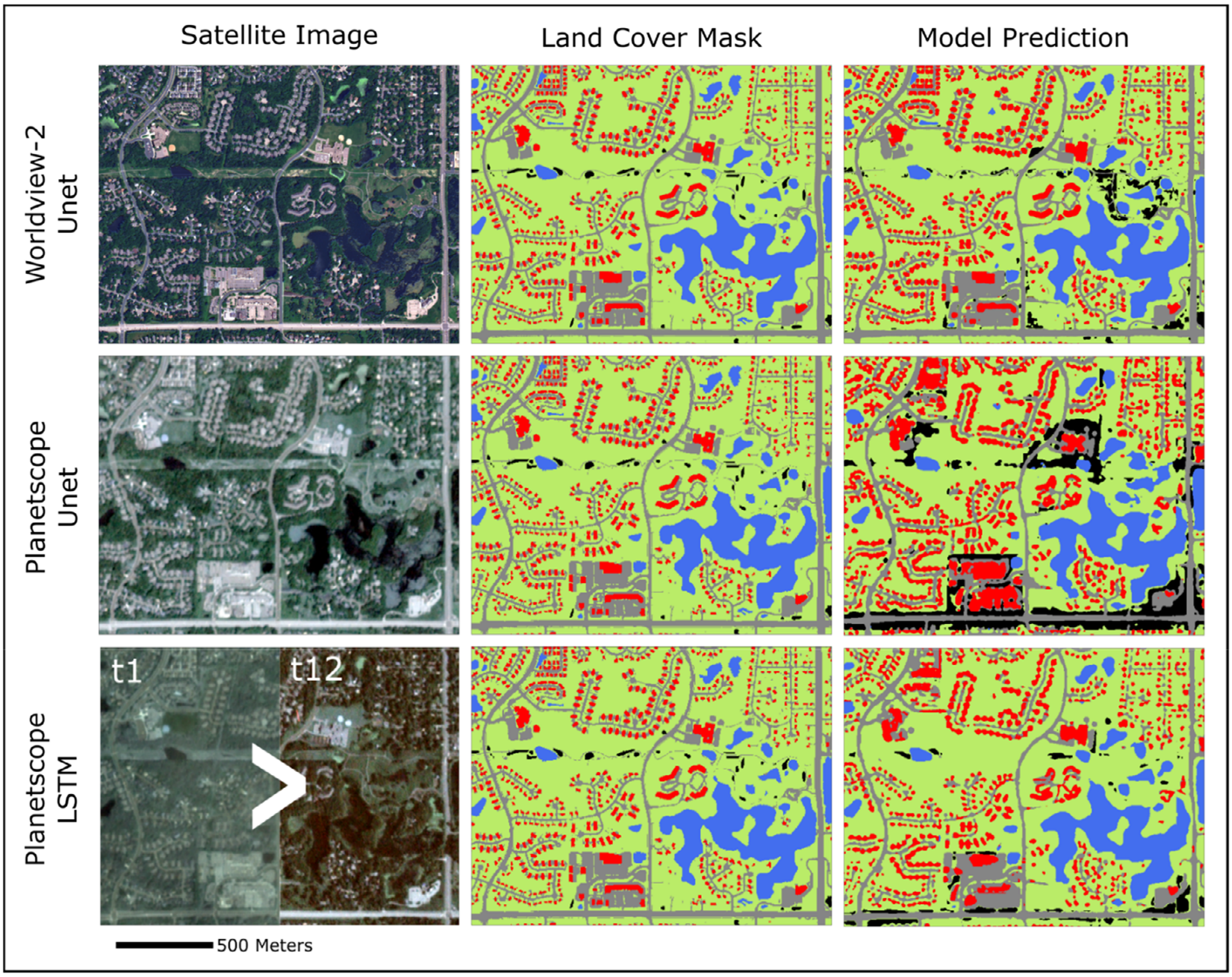

Figure 4. Sample of satellite images, land cover masks and model predictions from the Worldview-2 Unet (WV-CNN), Planetscope U-net (PS-CNN) and Planetscope LSTM (PS-LSTM) models. Top row: WV-CNN satellite image, land cover mask and predictions for leafy spurge (black), vegetation (green), roads (gray), buildings (red) and water (blue). Middle row: PS-CNN satellite image, land cover mask and predictions. Bottom row: PS-LSTM satellite images from first and last timestep, land cover mask and predictions.

Table 2. Planetscope U-net (PS-CNN) model evaluation metrics demonstrating model performance.

PS-CNN	Accuracy	Sensitivity	Specificity	Precision	F2 Score
Vegetation	0.777	0.635	0.94	0.924	0.678
Buildings	0.889	0.767	0.899	0.376	0.635
Roads	0.865	0.555	0.948	0.741	0.584
Water	0.963	0.945	0.965	0.749	0.898
Agriculture	0.956	0.897	0.96	0.616	0.822
Leafy Spurge	0.899	0.610	0.901	0.036	0.146

Table 3. Planetscope LSTM (PS-LSTM) model evaluation metrics demonstrating model performance.

PS-LSTM	Accuracy	Sensitivity	Specificity	Precision	F2 Score
Vegetation	0.830	0.780	0.890	0.898	0.801
Buildings	0.935	0.646	0.956	0.524	0.617
Roads	0.890	0.685	0.943	0.754	0.698
Water	0.977	0.953	0.980	0.855	0.931
Agriculture	0.966	0.928	0.968	0.602	0.837
Leafy Spurge	0.963	0.663	0.965	0.110	0.330

(96.6% vs. 95.6%, respectively), sensitivity (92.8% vs. 89.7%) and specificity (96.8% vs. 96.0%).

Deep learning model errors

The largest source of model error occurred due to the misclassification of leafy spurge and vegetation classes. We found that the majority of leafy spurge false positives (i.e. vegetation incorrectly predicted as leafy spurge) occurred within close proximity to leafy spurge true positives (Fig. 5). Approximately 51% of false-positive pixels were within 50 m (100 pixels) of a leafy spurge true positive for the WV-CNN model (Fig. 5). Comparatively, 35% of false-positive pixels were within 50 m for the PS-LSTM model, and 17% for the PS-CNN model (Fig. 5).

Model interpretation

We identified several internal layers from the WV-CNN and PS-CNN models that distinguish leafy spurge from other land cover classes (Fig. S5). However, visualizing individual layers provide limited understanding about which features drive overall predictions.

Using spectral manipulations of satellite imagery, we identified multiple spectral bands that aided in distinguishing leafy spurge from native vegetation. The WV-CNN model predictions were most sensitive to changes in the green, red-edge and near-infrared regions of the

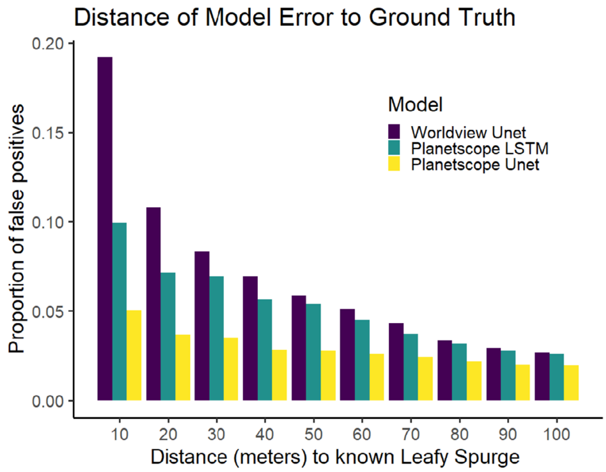

Figure 5. Distance between leafy spurge true positives and native vegetation false positives shown in purple for the Worldview-2 U-net (WV-CNN) model, yellow for the Planetscope U-net (PS-CNN) model, and teal for the Planetscope LSTM (PS-LSTM) model. The y-axis shows the proportion of false-positive pixels for native vegetation (i.e. vegetation predicted to be leafy spurge), and the x-axis displays distance between known leafy spurge true positives and false positives.

Worldview-2 imagery. Increasing or decreasing the mean of the sampled spectral reflectance for the green, red-edge or near-infrared band resulted in greater predicted probabilities for either leafy spurge or native vegetation (Figs. 3 and 6; Fig. S6). Similar patterns were identified when manipulating the green, red and near-infrared bands from Planetscope imagery (Fig. S7).

By manipulating phenological periods from the Planetscope LSTM model, we found that disrupting any of the time sets (4 images from early, mid or late season) reduced the predicted probability for leafy spurge, whereas it did not affect classes that lack a phenological trend (i.e. roads) (Fig. 7). Additionally, we found that disrupting images in the early season had the most dramatic impact on leafy spurge predictions. Using this approach for singular timesteps (i.e. a single image) did not appreciably change predictions of leafy spurge, indicating that the model was not reliant on a single timestep to make accurate predictions (results not shown). Together, these results indicate that the PS-LSTM model relies on information from several time periods (e.g. early and mid-season) that likely provide distinct signals necessary for detection.

Discussion

Recent advances in satellite imagery and deep learning architectures have provided a promising framework to detect and monitor individual species across large regions. In our study, we demonstrated that satellite image

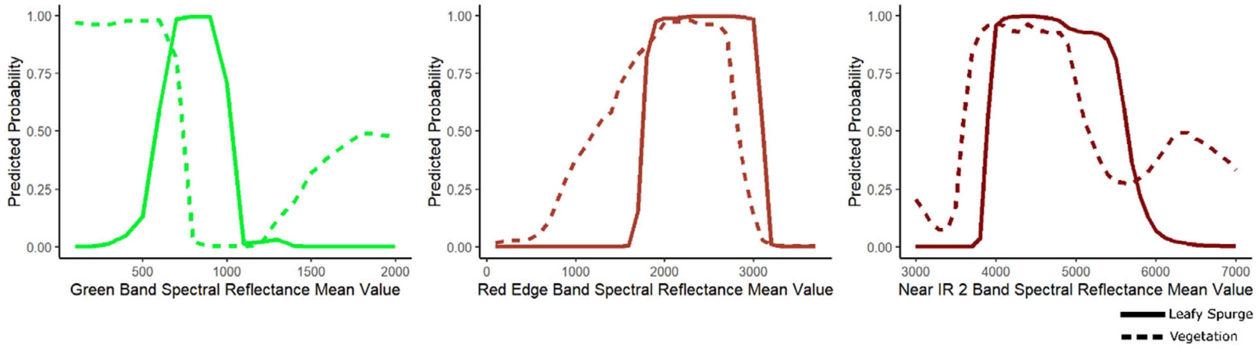

Figure 6. Manipulating spectral data show that changes in spectral information influence model predictions of leafy spurge and vegetation. Spectral data manipulation experiments with the green (left), red-edge (middle) and near-infrared-2 (right) spectral bands from Worldview-2 imagery. Spectral noise was first created by randomly sampling spectral bands from the leafy spurge and vegetation classes. Here, spectral noise was applied to Worldview-2 images and models were tasked to predict images containing noise. The mean of the spectral noise was incrementally increased across a regular interval to identify whether a model was sensitive to changes in spectral information. Increasing the mean value of the green spectral band lead to a greater predicted probability for leafy spurge, and decreasing the mean value of the green spectral band lead to a greater predicted probability for vegetation. Likewise, incrementally increasing the mean of the red edge and near-infrared 2 bands lead to shifts in predicted probabilities from the vegetation to leafy spurge class, indicating these spectral bands were likely important for leafy spurge prediction.

segmentation can accurately detect leafy spurge, an economically important invasive plant species, across complex landscapes. We found that U-net models constructed with high spatial and spectral resolution imagery (Worldview-2) had a higher performance for leafy spurge detection and land cover classification than those with lower spatial and spectral resolution (Planetscope). However, capturing growth and flowering phenology with an LSTM increased detection accuracy to the same level as with Worldview-2. These findings underscore the potential of using deep learning for geographic-scale plant detection and monitoring programs over complex landscapes.

One of our strongest findings was that incorporating phenology through a time series of images enhanced model performance. These results suggest that the relative timing of leafy spurge growth and flowering distinguishes it from background vegetation beyond its spectral signature. When we disrupted a fraction of the images (one third at a time), the prediction of leafy spurge was reduced but the prediction of other classes (e.g. roads) did not change (Fig. 7). Each time period that was disrupted (early, mid, late season) caused reductions in leafy spurge prediction; however, disruption of the early set of four images had the largest effect. The early set of four images includes the period where leafy spurge emerges from dormancy and the beginning of the flowering period. Because leafy spurge is among the earliest species to emerge in its habitat and also among the earliest to flower, this time frame appears to be particularly pivotal in distinguishing the species from background vegetation. Previous remote sensing studies of leafy spurge were only able to achieve high accuracy (>94%) using hyperspectral imagery and by distinguishing relative chlorophyll and carotenoid ratios at the red and near-infrared wavelengths (Glenn et al., 2005; Williams & Hunt, 2004). Our WV-CNN model similarly achieved high accuracy using submeter resolution, multispectral images that identified key spectral signatures in the green, red-edge and near-infrared bands. However, such high-resolution images are currently not available for most areas. As such, our study shows that the incorporation of temporal data can be a powerful tool for species detection even when high spatial and spectral resolution imagery is unavailable.

The growing availability of satellite imagery now allows researchers to detect and map plant species at regional scales (Kattenborn et al., 2021). While the WV-CNN model exhibited excellent performance at leafy spurge detection, its success depended on the availability of images from the peak flowering period. However, as Worldview-2 images are not taken at a regular spatial or temporal interval, they are likely to miss key timeframes and may fail to capture a species' true distribution and abundance (Müllerová et al., 2017). Thus, a time series of Planetscope images is advantageous for multiple reasons. At a local scale, relying on a single image to capture peak phenological signal requires local knowledge to determine when and where the phenomenon occurs. At regional scales, the timing of phenological signals (e.g. emergence, flowering and senescence) can differ significantly throughout a landscape due to spatial variability in climate and other factors (e.g. soil type) (Frantz et al., 2022; Weisberg

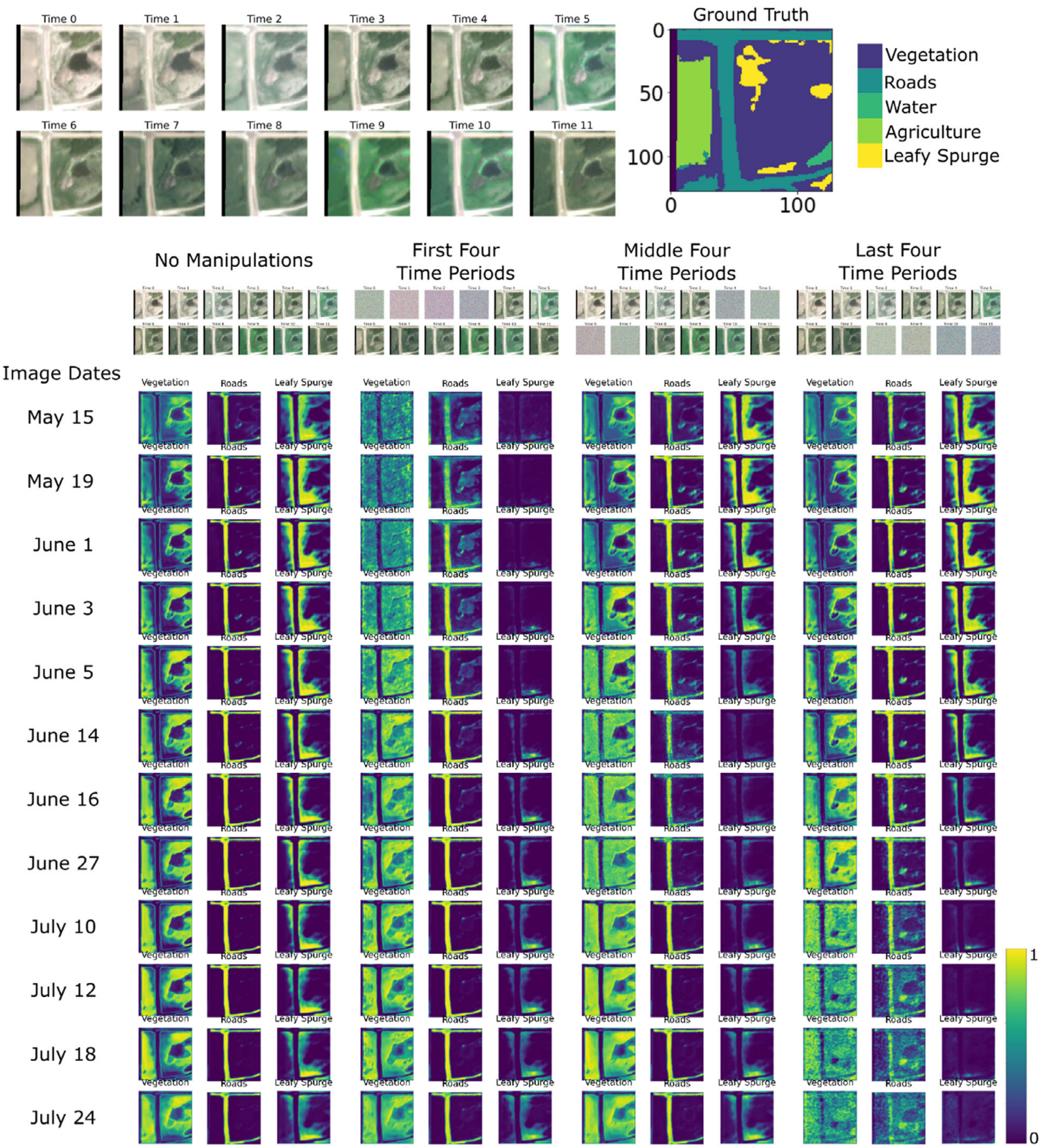

Figure 7. Satellite image manipulations designed to examine the importance of early season green-up, mid-season flowering and late season phenology on leafy spurge predictions from the Planetscope LSTM model. Top: the time series of 12 Planetscope images shows a typical phenological progression in the study area. Land cover map shows image labels, including vegetation, roads, water, agriculture and leafy spurge. Bottom: results of the time series image manipulations. After applying random Gaussian noise to images in the early season green-up (first four time periods), mid-season flowering (middle four time periods) and late season (last four time periods), comparing the predicted probability of vegetation, roads and leafy spurge classes shows that replacing any individual time periods reduces the probability for leafy spurge, and the early season green-up (first four time periods) had a substantial impact on leafy spurge predictions.

et al., 2021). In this case, a time series of images will be helpful to catch phenological signals in regions that are asynchronous. The region where leafy spurge has the highest impacts (Northern Great Plains of North America) contains high temporal and spatial variability in key phenological stages, making models based on single-date imagery highly dependent upon expert-based opinions for correctly timed imagery.

Successful image segmentation (i.e. pixel-wise image labeling) with deep learning relies on the delicate balance between minimizing both false-positive and false-negative errors. We demonstrated that most false-positive pixels between leafy spurge and native vegetation occurred close to surveyed populations. The close association between false-positive pixels and surveyed populations may occur because internal model layers detect leafy spurge and associated contextual habitat features (e.g. annual plants along roadsides), which may cause over-predictions of leafy spurge. Alternatively, the models may detect lower density patches at the margins of populations that we incorrectly digitized as other vegetation. In either case, the models predict areas that are consequential for invasive species management. False positives greater than 100 m from a documented leafy spurge population were often within inaccessible areas (e.g. airports). In these cases, we cannot discount the possibility that such predictions were indeed true populations.

Whereas the WV-CNN model detected most large (>1000 m^2) and small (<100 m^2) leafy spurge populations, the PS-LSTM model correctly identified many large populations (>1000 m^2) and resulted in few false positives where leafy spurge was absent (i.e. higher specificity) (Fig. S8). Despite lower sensitivity, the utility of Planetscope imagery for detecting and monitoring invasive plant species remains high. Temporal series of Planetscope imagery can provide natural resource managers or modelers with up-to-date information on a species' presence and abundance and detect landscape change through time (Curnick et al., 2021). Subsequently, information on invasive species can be used with biophysical datasets (e.g. climate and geological datasets) to generate habitat suitability maps (Jarnevich et al., 2021). Alternatively, if the goal of a remote sensing program is to detect small, nascent populations of invasive species, minimizing false-negative errors should be prioritized and higher resolution imagery should be strongly considered.

Accurate detection of plant species provides an avenue for tracking the distribution and abundance of a species in unprecedented ways. For example, remote sensing could be used to track the range expansion of invasive species, providing land managers with information on where to direct eradication efforts. For native species, individual detection would be valuable for demographic studies related to climate change when species are followed through time. Last, remotely sensed information on distribution and abundance are unbiased; whereas, publicly available databases on distributions suffer from the fact that observation effort is uneven across landscapes and through time. Such biases in occurrence records can be problematic in the development of species distribution models, which have been important for forecasting changes in species' ranges with climate change. Therefore, the application of deep learning to species-level detection has enormous potential to transform the long-term monitoring of distributions and the management of problematic species.

Acknowledgments

The funding for this project was provided by the Minnesota Invasive Terrestrial Plants and Pests Center through the Environment and Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR). We thank the Maxar Technologies and the European Space Agency for providing access to Worldview-2 satellite imagery (Proposal Id: 63324). We gratefully acknowledge the support of NVIDIA Corporation for the donation of the Titan V GPU used for this research. We thank Rob Venette and Naomi Schwartz for early discussions about ideas in this manuscript as well as Christian Klatt for records of leafy spurge. We also thank Lex Flagel for early review and discussions of deep learning.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J. et al. (2016) TensorFlow: a system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*. Berkeley: USENIX Association, pp. 265–283.

Asner, G.P. & Martin, R.E. (2008) Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. *Remote Sensing of Environment*, **112**, 3958–3970.

Bangsund, D.A., Leistritz, F.L. & Leitch, J.A. (1999) Assessing economic impacts of biological control of weeds: the case of leafy spurge in the northern Great Plains of the United States. *Journal of Environmental Management*, **56**, 35–43.

Belcher, E.A. & Wilson, S.D. (1989) Leafy spurge and the species composition of mixed-grass prairie. *Journal of Range Management*, **42**, 172–175.

Bradley, B.A. (2014) Remote detection of invasive plants: a review of spectral, textural and phenological approaches. *Biological Invasions*, **16**, 1411–1425.

Bradley, B.A., Laginhas, B.B., Whitlock, R., Allen, J.M., Bates, A.E., Bernatchez, G. et al. (2019) Disentangling the abundance-impact relationship for invasive species.

Proceedings of the National Academy of Sciences of the United States of America, **116**, 9919–9924.

Cabezas, M., Kentsch, S., Tomhave, L., Gross, J., Caceres, M.L.L. & Diez, Y. (2020) Detection of invasive species in wetlands: practical DL with heavily imbalanced data. *Remote Sensing*, **12**, 3431.

Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G. et al. (2020) Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific Reports*, **10**, 17188.

Chollet, F. (2015) Keras, GitHub repository. Available from: https://github.com/fchollet/keras [Accessed 8th December 2020].

Clark, M.L., Roberts, D.A. & Clark, D.B. (2005) Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, **96**, 375–398.

Clinton, N.E., Potter, C., Crabtree, B., Genovese, V., Gross, P. & Gong, P. (2010) Remote sensing-based time-series analysis of cheatgrass (*Bromus tectorum* L.) phenology. *Journal of Environmental Quality*, **39**, 955–963.

Curnick, D.J., Davies, A.J., Duncan, C., Freeman, R., Jacoby, D.M.P., Shelley, H.T.E. et al. (2021) SmallSats: a new technological frontier in ecology and conservation? *Remote Sensing in Ecology and Conservation*, **8**, 139–150. https://doi.org/10.1002/rse2.239

Dahal, D., Pastick, N.J., Boyte, S.P., Parajuli, S., Oimoen, M.J. & Megard, L.J. (2022) Multi-species inference of exotic annual and native perennial grasses in rangelands of the Western United States using harmonized landsat and sentinel-2 data. *Remote Sensing*, **14**, 807.

Diagne, C., Leroy, B., Vaissière, A.-C., Gozlan, R.E., Roiz, D., Jarić, I. et al. (2021) High and rising economic costs of biological invasions worldwide. *Nature*, **592**, 571–576.

DiTomaso, J.M. (2000) Invasive weeds in rangelands: species, impacts, and management. *Weed Science*, **48**, 255–265.

Duncan, C.A., Jachetta, J.J., Brown, M.L., Carrithers, V.F., Clark, J.K., DiTomaso, J.M. et al. (2004) Assessing the economic, environmental, and societal losses from invasive plants on rangeland and wildlands. *Weed Technology*, **18**, 1411–1416.

Dunn, P.H. (1979) The distribution of leafy spurge (*Euphorbia esula*) and other weedy *Euphorbia* spp. n the United States. *Weed Science*, **27**, 509–516.

Dymond, C.C., Mladenoff, D.J. & Radeloff, V.C. (2002) Phenological differences in tasseled cap indices improve deciduous forest classification. *Remote Sensing of Environment*, **80**, 460–472.

ESRI. (2020) *ArcGIS desktop: release 10*. Redlands: Environmental Systems Research Institute.

Frantz, D., Hostert, P., Rufin, P., Ernst, S., Röder, A. & van der Linden, S. (2022) Revisiting the past: replicability of a historic long-term vegetation dynamics assessment in the era of big data analytics. *Remote Sensing*, **14**, 597.

Frazier, A.E. & Hemingway, B.L. (2021) A technical review of planet smallsat data: practical considerations for processing and using planetscope imagery. *Remote Sensing*, **13**, 3930.

Gaskin, J.F., Espeland, E., Johnson, C.D., Larson, D.L., Mangold, J.M., McGee, R.A. et al. (2020) Managing invasive plants on Great Plains grasslands: a discussion of current challenges. *Rangeland Ecology and Management*, **78**, 235–249.

Ghosh, R., Ravirathinam, P., Jia, X., Khandelwal, A., Mulla, D., Kumar, V. (2021). *CalCROP21: a georeferenced multi-spectral dataset of satellite imagery and crop labels*. arXiv:2107.12499.

Glenn, N.F., Mundt, J.T., Weber, K.T., Prather, T.S., Lass, L.W. & Pettingill, J. (2005) Hyperspectral data processing for repeat detection of small infestations of leafy spurge. *Remote Sensing of Environment*, **95**, 399–412.

Goodfellow, I., Bengio, Y. & Courville, A. (2016) *Deep learning*. Cambridge: MIT press.

Graves, A., Fernández, S. & Schmidhuber, J. (2005) Bidirectional LSTM networks for improved phoneme classification and recognition. In: Duch, W., Kacprzyk, J., Oja, E. (Eds.) *Artificial neural networks: formal models and their applications – ICANN 2005*. Berlin Heidelberg: Springer, pp. 799–804.

He, K., Zhang, X., Ren, S., Sun, J. (2015) *Deep residual learning for image recognition*. arXiv:1512.03385.

Hochreiter, S. & Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, **9**, 1735–1780.

Hooge Hom, K.M., Bajwa, S.G., Lym, R.G. & Nowatzki, J.F. (2020) Discrimination of leafy spurge (*Euphorbia esula*) and purple loosestrife (*Lythrum salicaria*) based on field spectral data. *Weed Technology*, **34**, 250–259.

Host, T.K.; Rampi, L.P & Knight, J.F. (2016) *Twin cities metropolitan area 1-meter land cover classification (impervious surface focused)*. Retrieved from the Data Repository for the University of Minnesota. https://doi.org/10.13020/D6959B.

Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. (2020) Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2011–2023.

Huang, C.Y. & Asner, G.P. (2009) Applications of remote sensing to alien invasive plant studies. *Sensors*, **9**, 4869–4889.

Hunt, E.R., McMurtrey, J.E., Williams, A.E.P. & Corp, L.A. (2004) Spectral characteristics of leafy spurge (*Euphorbia esula*) leaves and flower bracts. *Weed Science*, **52**, 492–497.

Irons, J.R. & Petersen, G.W. (1981) Texture transforms of remote sensing data. *Remote Sensing of Environment*, **11**, 359–370.

Isaacson, B.N., Serbin, S.P. & Townsend, P.A. (2012) Detection of relative differences in phenology of forest species using Landsat and MODIS. *Landscape Ecology*, **27**, 529–543.

Jarnevich, C.S., Sofaer, H.R. & Engelstad, P. (2021) Modelling presence versus abundance for invasive species risk assessment. *Diversity and Distribution*. https://doi.org/10.1111/ddi.13414

Kattenborn, T., Eichel, J., Wiser, S., Burrows, L., Fassnacht, F.E. & Schmidtlein, S. (2020) Convolutional neural networks accurately predict cover fractions of plant species and communities in unmanned aerial vehicle imagery. *Remote Sensing in Ecology and Conservation*, **6**, 472–486.

Kattenborn, T., Leitloff, J., Schiefer, F. & Hinz, S. (2021) Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, **173**, 24–49.

Kattenborn, T., Lopatin, J., Förster, M., Braun, A.C. & Fassnacht, F.E. (2019) UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, **227**, 61–73.

Kingma, D.P. & Ba, J. (2014) Adam: a method for stochastic optimization. *arXiv* preprint, arXiv:1412.6980.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2017) ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **60**, 84–90.

Larson, K.B. & Tuor, A.R. (2021) Deep learning classification of cheatgrass invasion in the western United States using biophysical and remote sensing data. *Remote Sensing*, **13**, 1246.

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, **521**, 436–444.

Leistritz, F.L., Thompson, F. & Leitch, J.A. (1992) Economic impact of leafy spurge (*Euphorbia esula*) in North Dakota. *Weed Science*, **40**, 275–280.

Leitch, J.A., Leistritz, F.L. & Bangsund, D.A. (1994) Economic effect of leafy spurge in the Upper Great Plains: methods, models, and results. *Impact Assessment*, **14**, 419–433.

Levin, G.A. & Gillespie, L.J., 2017. Euphorbiaceae. In: Flora of North America Editorial Committee (Eds.). *Flora of North America*. New York: Oxford. Available at: http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=10327. [Accessed 27th October 2021].

Lym, R.G. (2005) Integration of biological control agents with other weed management technologies: successes from the leafy spurge (*Euphorbia esula*) IPM program. *Biological Control*, **35**, 366–375.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. & Johnson, B.A. (2019) Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **152**, 166–177.

Marbuah, G., Gren, I.-M. & McKie, B. (2014) Economics of harmful invasive species: a review. *Diversity*, **6**, 500–523.

Martin, M.E., Newman, S.D., Aber, J.D. & Congalton, R.G. (1998) Determining forest species composition using high spectral resolution remote sensing rata. *Remote Sensing of Environment*, **65**, 249–254.

Mäyrä, J., Keski-Sarri, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P. et al. (2021) Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, **256**, 112322.

Mazzia, V., Khaliq, A. & Chiaberge, M. (2020) Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-convolutional neural network (R-CNN). *Applied Sciences*, **10**, 238.

Müllerová, J., Brůna, J., Bartaloš, T., Dvořák, P., Vítková, M. & Pyšek, P. (2017) Timing is important: unmanned aircraft vs. satellite imagery in plant invasion monitoring. *Frontiers in Plant Science*, **8**, 887.

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H. & Turner, W. (2014) Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, **51**, 839–848.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Planet Labs, Inc. (2020) *Planet planet imagery product specifications—June 2020*. San Francisco: Planet Labs, Inc.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. et al. (2019) Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204.

Rist, Y., Shendryk, I., Diakogiannis, F. & Levick, S. (2019) Weed mapping using very high resolution satellite imagery and fully convolutional neural network. In: *IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 9784–9787.

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

Roy, D.P., Huang, H., Houborg, R. & Martins, V.S. (2021) A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. *Remote Sensing of Environment*, **264**, 112586.

Rußwurm, M. & Körner, M. (2018a) Multi-temporal land cover classification with sequential recurrent encoders. *International Society for Photogrammetry and Remote Sensing*, **7**, 129.

Rußwurm, M., Körner, M. (2018b) *Convolutional LSTMs for cloud-robust segmentation of remote sensing imagery*. arXiv:1811.02471.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K. & Müller, K.-R. (Eds.). (2019) *Explainable AI: interpreting, explaining and visualizing deep learning*. Cham: Springer.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B. et al. (2020) Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, **170**, 205–215.

Schmidt, J.P., Springborn, M. & Drake, J.M. (2012) Bioeconomic forecasting of invasive species by ecological syndrome. *Ecosphere*, **3**, 46.

Schuster, M. & Paliwal, K.K. (1997) Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, **45**, 2673–2681.

Seebens, H., Blackburn, T.M., Dyer, E.E., Genovesi, P., Hulme, P.E., Jeschke, J.M., et al. (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* **8**, 14435.

Sefrin, O., Riese, F.M. & Keller, S. (2020) Deep learning for land cover change detection. *Remote Sensing*, **13**, 78.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. & Woo, W.-C. (2015). *Convolutional LSTM network: a machine learning approach for precipitation nowcasting*. arXiv: 1506.04214.

Shorten, C. & Khoshgoftaar, T.M. (2019) A survey on image data augmentation for deep learning. *Journal of Big Data*, **6**, 1–48.

Sofaer, H.R., Jarnevich, C.S. & Pearse, I.S. (2018) The relationship between invader abundance and impact. *Ecosphere*, **9**, e02415.

Wallace, N.M., Leitch, J.A., Larry Leistritz, F. (1992) *Economic impact of leafy spurge on North Dakota wildland*. NDSU Agricultural Economics Report No. 281.

Weisberg, P.J., Dilts, T.E., Greenberg, J.A., Johnson, K.N., Pai, H., Sladek, C. et al. (2021) Phenology-based classification of invasive annual grasses to the species level. *Remote Sensing of Environment*, **263**, 112568.

Williams, P.A.E. & Hunt, R.E. (2004) Accuracy assessment for detection of leafy spurge with hyperspectral imagery. *Journal of Range Management*, **57**, 106–112.

Wolter, P.T., Mladenoff, D.J., Host, G.E. & Crow, T.R. (1995) Improved forest classification in the northern lake states using multi-temporal Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, **61**, 1129–1143.

Xue, J. & Su, B. (2017) Significant remote sensing vegetation indices: a review of developments and applications. *Journal of Sensors*, **2017**, 1353691.

Yakubovskiy, P. (2019) Segmentation models, GitHub repository. Available from: https://github.com/qubvel/segmentation_models [Accessed 1st June 2021].

Yu, Y., Si, X., Hu, C. & Zhang, J. (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, **31**, 1235–1270.

Supporting Information

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

Table S1. Number of 1-m squared pixels in the ground truth raster for each land cover class present within six areas of interest (AOIs) in study area.
Methods S1. Methods for Worldview-2 satellite image corrections and preprocessing.
Table S2. Worldview-2 imagery metadata.
Table S3. Planetscope imagery metadata.
Figure S1. Spectral discrimination of leafy spurge from co-occurring vegetation.
Figure S2. Representative sample of 1-m resolution land cover map used for the land cover mask during deep learning model training.
Figure S3. Diagram of the U-net for satellite image segmentation.
Figure S4. ConvLSTM layer.
Table S4. Worldview-2 and Planetscope CNN and LSTM model error types.
Figure S5. Visualizing internal U-net model layers.
Figure S6. Results of spectral data manipulation using Worldview-2 satellite imagery.
Figure S7. Results of spectral data manipulation using Planetscope satellite imagery.
Figure S8. Segmentation models predict larger leafy spurge populations with greater frequency.

