Title

Integrating Deep Learning and Google Street View for Novel Weed Mapping

Author

Zhen, Tong

Publication Date

2022

Peer reviewed|Thesis/dissertation

Integrating Deep Learning and Google Street View for Novel Weed Mapping

By

TONG ZHEN

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Horticulture and Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Mohsen B. Mesgaran, Chair

Kassim Al-Khatib

Daniel E. Runcie

Committee in Charge

2022

Abstract

Road facilitates invasive plant species' early establishment and spread, which can impact ecosystem services and cause economic loss. Monitoring invasive species populations along the road is important in roadside integrated vegetation management (IVM). A roadside weedy and invasive species map can assist in developing species population models and designing proper weed management strategies. However, the traditional road survey requires unrealistic human labor and time to map invasive species at large scales. A novel weed mapping system was developed to retrieve species location data by integrating Google Street View (GSV) imagery and object detection based on deep learning algorithms. The target species of this feasibility study was johnsongrass (*Sorghum halepense*). We trained the detection network, You Only Look Once (YOLOv2), with 911 johnsongrass roadside images retrieved from Google Street View. YOLOv2 is a fast and accurate deep neural network. The trained detection model could detect johnsongrass in GSV images, and output bounding boxes with the target species' confidence scores. The trained model was then applied to a large image dataset of ~270,000 images along 135,000 km of roads in California, Oregon, Washington, and Nevada. The network detected 2,031 new johnsongrass records along roads in these four states, and the location of each image was used to create a map of the johnsongrass population. Our current deep learning model has 85% recall, 73.9% precision, and 77.5% overall accuracy on the testing dataset, which included 2,040 images. The model also has a 30% false positive rate (FPR). Work is in progress to reduce the FPR. Using our novel AI-based method, the estimated cost of the weed survey in four states is $3,570, while the traditional road surveys with cars at the same

scale will cost at least $63,558 without considering associated risks, such as car accidents. Besides that, traditional road surveys require six months, but the automated weed survey only requires a few days with a trained detection network. The automated mapping scheme can apply to other weedy and invasive species, and it is possible to map this weed (and others) on a much larger scale, which is the focus of our future work.

Key words: Google Street View, deep learning, weed survey, *Sorghum halepense*, integrated vegetation management, IVM, invasive species, roadside weed management

Acknowledgements

Firstly, I want to thank everyone who helps me succeed academically in this master's program. I really appreciate my major professor, Dr. Mohsen Mesgaran, for providing such a good experience and project in his lab. I also want to thank Dr. Kassim Al-Khatib and USDA for funding this project. This thesis work will not be better without the help of my thesis committee members, including Dr. Mohsen Mesgaran, Dr. Kassim Al-Khatib, and Dr. Daniel Runice. I also want to thank my academic advisor, Dr. Brad Hanson, for his support and advice over the past three years.

Secondly, I want to say thank you to my mother, Liping He, and my family back in China. My mother always tries her best to give me a better life and a better education opportunity. Then I want to say thank you to my significant one, Chen Zhong. You always support me and stand by my side. Lastly, I want to say sorry to my uncle, Jiajun Zhen, and my grandfather, Zhuoming Zhen. I will never have a chance to talk to you both anymore, and I really hope that you both can have the opportunity to read my work and say, "I am really proud of you." You both always told me to be a good man and be nice to everyone around me. My grandfather is a wise man and has been my role model since the very first day. My uncle loves reading and sciences, and he is the one who inspires me to do research. I wouldn't have this accomplishment today without all the people I mentioned above, and I really love you all.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Roadside Weed Species

Since the 1930s, the growth of road transportation has exploded in the United States, and about 83% of the land in the United States is within one kilometer of any type of road (Riitters & Wickham, 2003; Taaffe et al., 1996, pp. 115-116). Road infrastructure facilitates economic growth and human-social interaction; however, it is also well-known that roads facilitate invasive and weedy plant dispersal (Jodoin et al., 2008; Ansong & Pickering, 2013). In most cases, roads act as barriers and filters to block most wildlife species' movement because of the frequent and intense disturbances; however, roads also act as habitats and conduits for invasive and weedy species adapted to the disturbances (Ree et al., 2015, p. 6). Plants along the roads usually raise safety concerns, such as blocking traffic and affecting drivers' vision. Additionally, massive road networks have now become a part of the ecosystem. Since roads connect both natural landscapes and agricultural fields, roadside vegetation management is important to minimize the impacts of invasive and weedy species on those ecosystems.

In 2021, a study demonstrated that invasive species, including plants and animals, cost North America $1.26 trillion from 1960 to 2017 and $26 billion annually in the 2010s (Crystal-Ornelas *et al*., 2021). In 2008, according to the survey by California Invasive Species Council, the economic impact of invasive plant species was estimated at $82 million annually (California Invasive Plant Council, 2017). In addition to the economic impact and safety concerns, invasive and weedy plants can cause ecological damages to the ecosystem, including reduction of species biodiversity, changes in

wildfire regime, and water pollution (Das & Duarah, 2013; Gelbard & Belnap, 2003). A changing fire regime can affect local species and human society to a large extent, so managing roadside invasive species is inevitable and necessary. For example, *Bromus tectorum* and *Andropogon gayanu* can increase fire intensity and frequency by adding more fuel load to the ecosystem (Brooks *et al*., 2004).

In general, a successful plant species establishment consists of several factors. Reichard & Hamilton (1997) suggested that weedy traits, especially reproductive traits, are the most dominant factors in determining a successful invasion of an ecosystem. In contrast, Parendes & Jones (2000) argued that environmental factors (nutrient availability and disturbance) or human interferences also partially explained the invasive species' distribution and dispersal. In the case of roadside habitats, human activities and the environment are also as prominent as weedy traits. The frequent disturbances along the road, including road maintenance and building, create long-lasting bare soil for species colonization. Parendes & Jones (2000) reported that locations with intensive disturbances and adequate resources (e.g., sunlight exposure and nutrients) have higher frequencies of exotic plant species. Frequent and intense disturbances can change soil chemical and physical properties. Mills *et al*. (2021) measured and examined the soil properties in two segments from two different highways in Nebraska. The data indicated that roadside soil contained high sodium concentration and high soil compaction, which inhibited the growth of native vegetation (Mills *et al*., 2021). A meta-analysis demonstrated that invasive species have higher plasticity than co-occurring native species under many environmental stresses (Davidson *et al*., 2011). Besides disturbances, water resources are another factor that increases the invasibility of

roadside environments. Roads are considered water collectors and rainfall storage. The surface of roads is designed to have a few degrees of incline to drain excess water during rainfall, and there will be a slide slope with more incline and a ditch to collect the water (Bohemen & Janssen van de Laak, 2003). A study measured the soil moisture of the ditch along forest roads in 36 samples, averaging 53.5% (Neher *et al*., 2013). Adequate water resources along the road provide a suitable environment and resources for seed germination and early seedling growth for invasive species. After the early establishment and the naturalized species overcoming different stresses to produce seeds, the rapid spread of the reproductive offspring makes the species invasive (Richardson *et al*., 2000).

Human-assisted dispersal potentially creates a longer-range spread than the dispersal mechanisms related to species' reproductive traits. According to Mortensen *et al*. (2009), human activities are the main facilitators of the weedy and invasive species spread, and their study indicated that paved roads could spread more weeds than forests and wetlands. Specifically, vehicles are the spreading vectors of long-distance dispersal (LDD) for weedy and invasive species. The traditional species dispersal model described LDD as a rare event; most cases are seed dispersal by animals (epizoochory), where weed seeds are adhesive to animal fur and travel along with seasonal migration (Loebach & Anderson, 2018). However, Nathan (2003) proposed that human-mediated LDD has become the most important mechanism of LDD in plants and animals, which is a challenge for future LDD prediction. According to Baker (1974), no specific weedy traits or natural dispersal mechanism to help invasive species overcome the large-scale geographical barrier. Nevertheless, human-assisted dispersal

can potentially transfer seeds over 100 km away from the parent plants. The 100 km was an approximated cut-off to classify plants as alien species in the model proposed by Richardson *et al*. (2000). A study in Germany collected seeds from roadside verges, and their results indicated that nearly 30% of the species collected were by LDD and some species they identified are highly invasive in other countries (von der Lippe & Kowarik, 2007). Seeds dispersed by vehicles share common characteristics that might facilitate car-borne dispersal. Zwaenepoel *et al*. (2006) provided another perspective by collecting seed samples from mud attached to the car; the results suggested that car-borne floras were pioneer species with small and light seeds. Other traits like large seed production and the ability to reproduce vegetatively are also reported in many studies (Ansong & Pickering, 2013). Also, a systematic review summarized that about 626 species in 75 families had been identified from cars, and Poaceae is the most frequent family (28%), followed by Asteraceae (13%) and Fabaceae (7%) (Ansong & Pickering, 2013).

The spread of invasive and weedy species along the road in the real world could be more significant than what has been reported in scientific studies. The large-scale spread of *Microstegium vimineum* was reported by observation; however, according to a model prediction, the species spread only by natural dispersal is limited (Rauschert *et al*., 2009). The contrasting results support that human-assisted dispersal leads to an increasing spread rate, and more resources and efforts should be put to roadside vegetation management.

Roadside vegetation management is on a large scale, and the local government is the primary agent for management. For example, in California, the California

Department of Transportation (Caltrans) manages the vegetation along California state highways, and develops projects to protect motorists, cyclists, and potential wildfire spread along the roads (California Department of Transportation, n.d.). Compared to agricultural weed management, roadside vegetation management has limited tools. The most common practice is mowing, but it requires multiple applications in a short period; therefore, mowing is expensive and ineffective since only the foliar part of the plant is damaged (Hyman & Vary, 1999). Herbicide application is used with mowing as the Integrated Vegetation Management (IVM). Chemical control could be effective under roadside conditions. For example, herbicide trials conducted in six different regions of Indiana demonstrated that herbicide application could effectively control broadleaf species for more than one year and grasses for months (Herold *et al*., 2014). Herbicides can significantly lower the cost, but the increasing herbicide-resistant population is another potential concern for roadside weed management (Bagavathiannan & Norsworthy, 2016). IVM program is important to manage roadside invasive species, and similar to Integrated Weed Management, early detection and monitoring are also main components in IVM. The first step in studying and analyzing the ecological aspect of a specific invasive or weedy species is to conduct a species survey and map the population distribution.

1.2. Methods to Create Species Map

A species distribution map is a common approach for evaluating the extent of plant invasion and provides a baseline for the informed allocation of resources and efforts. Botanists usually conduct field surveys to collect plant species, including weedy

species. The benefit of this detailed survey is the high accuracy of the species identification and location data, but a detailed survey requires enormous resources. For example, a county-level survey of 3000 km required 35 months, and the researcher had to travel by car, on foot, or even by boat (Abella *et al*., 2008). A typical 3000 km survey is considered small-scale but still time-consuming, labor-intensive, and requires equipment like a vehicle and an accurate GPS positioning system. A field survey is reasonable for species local population examination and species-environment interaction analysis.

A car survey can be rapid by applying different sampling or examination methods. According to Shuster *et al*. (2005), a car survey has a similar probability of finding *Alliaria petiolata* compared to a survey on foot but requiring four times fewer person-hours. The car survey can involve transects or random sampling sites along the roads base on land uses, soil types, rainfall, and vegetation types (Milton & Dean, 1998; Shuster *et al*., 2005; Ohadi *et al*., 2018). The data collected from the car survey can be used to build a model to understand the relationship between species distribution and environmental factors. A car survey could yield consistent results when various factors are examined in the experimental design. For instance, the traveling speed can vary for different types of roads. For highways, the driver must drive above the minimum speed so that a higher speed can result in lower identification accuracy. Observation accuracy is another factor that affects data consistency. Catry *et al*. (2015) conducted accuracy tests to evaluate the potential human errors. However, it is challenging to include human error rates in the species distribution maps, and in most cases, the human errors in the car survey are unaccountable (Milton & Dean, 1998; Shuster *et al*., 2005). A

standard methodology for roadside species surveys should be established to yield consistent and comparable data.

Some studies do not include all road types in the car survey to save time and reduce costs. For example, Catry *et al*. (2015) excluded the highway or freeway because they believed that the roadsides of the freeway are well-managed and there will be less possibility of having invasive species populations. A survey for all roads in a state will take an unrealistic time to complete, and the cost of traveling will be expensive. For example, California has about 622,000 km of roads, and it will take about 780 days if a driver drives 800 km per day (Federal Highway Administration, n.d.). Thus, government agencies will hesitate to conduct surveys because of limited funding and resources. However, roadside vegetation assessment can help identify the level of invasion and the potential damage. A car survey is not cost-effective for accomplishing a quick assessment on a large scale.

Furthermore, a species map can be used in large-scale species dispersal models in which the input data are usually from global or regional databases. Kadmon *et al*. (2004) argued that since randomized surveys on a large scale are rare, those models often rely on incomplete databases with biased data, such as herbaria, natural museums, and user-uploaded entries. The unified database includes data collected from different observers, and we cannot estimate the potential human error if we rely on these databases to run the ecological models. As a result, we need a more systematic approach for large-scale species surveys.

1.3. Google Street View

Google Street View (GSV) is a tool in Google Maps and Google Earth that allows users to interact with the panoramas along streets and roads in many countries. GSV and Google Earth are well-developed and well-maintained databases, and the imagery has been online for more than ten years. Google sends numerous data-collection vehicles on the roads, and those vehicles take 360-degree photos, e.g., by installing a rosette camera on the top of the car (Anguelov *et al*., 2010). According to Anguelov *et al*. (2010), this project aims to organize a large amount of information, and billions of users can have access to that information. GSV is well-known among ordinary users for educational and recreational activities, but now, these images can be used for ecological studies and vegetation management. For example, GSV was used to map the distribution of the Pine Processionary Moth (*Thaumetopoea pityocampa*), in an area of about 45,000 km^2 (Rousselet *et al*., 2013). Their research suggested that this method can be effective if the target species are distinguishable in the GSV images. They also mentioned that the coverage area of GSV still needed to be completed back in 2013.

In the past ten years, GSV has been used in several ecological studies. For example, according to Hardion *et al*. (2016), the integration of ground and aerial images could create a better species distribution map of giant cane (*Arundo donax L.*), a common grass species along the road, and the species distribution data produced better results in species distribution models compared to the traditional field survey. Additionally, two studies compared GSV and field surveys by car. Studies by Deus *et al*. (2016) and Kotowska *et al*. (2021) reported that the results produced by GSV resemble that produced by the field survey. The two studies surveyed different species of different

sizes (trees and flowering plants). The results indicated the versatility and high performance of the GSV database. Although GSV is more cost-effective than car surveys, studies discussed above have used human observers for plant detection, which is tedious, time-consuming, and not feasible for large-scale mapping.

Furthermore, most previous studies reported some common limitations of this imagery database. Most studies suggested seasonality and time differences among the GSV images. The images from different areas were taken in different years or seasons and at different times during the day (morning to evening), which could impact the presence and absence of the survey species or detectability of species within the GSV images (Rousselet *et al*., 2013; Hardion *et al*., 2016; Deus *et al*., 2016; Kotowska *et al*., 2021). Deus *et al*. (2016) also reported that contrast, ambient light, and sharpness would affect the identification accuracy in some images. Another issue with GSV is that most species were undetectable in their seedling stage when they were small and did not develop distinguishable features (flowers, leaf shape, and plant structure) (Kotowska *et al*., 2021).

GSV database has been successfully used as a cost-effective method in terms of time and resources, and this benefit can allow researchers to conduct a comparatively large-scale survey in a short period (Rousselet *et al*., 2013; Visser *et al*., 2013; Hardion *et al*., 2016; Deus *et al*., 2016; Kotowska *et al*., 2021).

1.4. Computer Vision

Artificial Intelligence (AI) is a powerful tool to automate tasks, providing insightful solutions for scientific studies in different fields. For example, image detection built with

deep learning algorithms can be used in plant species identification. Computer vision methods are successfully used in crop and weed classification to build robotic machines to conduct real-time weed detection in the field (Wang *et al*., 2019).

Machine learning includes two main types: unsupervised and supervised learning. Unsupervised learning trains models with unlabeled data, while supervised learning requires labeled data, and most neural networks are examples of supervised learning algorithms (Géron, 2019). Artificial neural networks (ANN) are models built based on the nervous system of the living vertebrate (Fukushima, 1980). In recent years, neural networks have evolved into the convolutional neural network (CNN), which consists of multiple convolutional layers.

Many studies have proven that CNNs can reach high accuracy in object detection in images (Dang *et al*., 2022; Dyrmann *et al*., 2016; Yan & Ryu, 2021). Dyrmann *et al*. (2016) trained a CNN model to classify young seedlings (2 to 10 days old) of 22 species with the average accuracy of 86.2% under controlled indoor conditions. For example, Sugar beet (*Beta vulgaris*), barley (*Hordeum vulgare L.*), and Thale Cress (*Arabidopsis thaliana*) could achieve 97% to 98% accuracy (Dyrmann *et al*., 2016). Dyrmann *et al*. (2017) tested another CNN model called DetectNet in overhead images from highly occluded cereal fields, and the model had a recall of 46.3% and a precision of 86.6% for detecting the non-crop plants. The high-occluded fields resemble roadside environments where plants overlap each other.

Another well-known fast-detection CNN model called You Only Look Once (YOLO) can be used to detect weeds in outdoor and natural light conditions. Dang *et al*. (2022) tested different versions of YOLO detectors on 12 weed classes at different

growth stages in cotton fields. The precisions of 12 weed classes ranged from 81.5% to 98.28%, and the recalls ranged from 78.62% to 97.9% on the YOLOv3 detector (Dang *et al*., 2022). The examples above are all based on overhead images in crop fields, but only a few similar studies have been done under roadside conditions. One example of the roadside condition is a study that integrated GSV Imagery and a CNN network to map the distribution of different crops along the roads (e.g., alfalfa, almonds, corn) and achieved accuracy levels of 92% in California and 98% in Illinois (Yan & Ryu, 2021). Agronomic crops have distinct and uniform morphology, but for roadside invasive species, the high variation in plant morphology and the non-uniform backgrounds (e.g., different vegetation conditions) will lead to more significant detection errors.

Previous studies and discussions supported that computer vision can replace human observers in species detection. However, we still need human observers to create a training dataset. Training, testing, and validation datasets are the essential components of a deep learning model, with the most time-consuming task being image labeling. A larger training dataset can increase model performance, but the size would determine the amount of labor for a single project. Abdulsalam & Aouf (2020) suggested that 1,000 images of a particular species are required to achieve high prediction accuracy. Yan & Ryu (2021) proposed that the training sizes would differ depending on the mapping species since they only used 400 training samples for corn, but the model could still perform with high accuracy. Compared to the traditional car survey, the AI-based survey method can be conducted by non-experts once the image detection model is trained. GSV and image detection algorithms can perform large-scale weed mapping with low resource input.

1.5. Johnsongrass

Johnsongrass is a common weed along the roads in the United States and is a good model plant for the AI-based survey method. It is native to the Mediterranean and North Africa and was introduced to the United States in the early 1800s (Oyer *et al*., 1959; Bhatti *et al*., 1960). Johnsongrass is a perennial grass and can colonize and spread nearby landscapes through the underground rhizome system (Oyer *et al*., 1959). The mature plant can grow up to 2.5 meters in height, and the height of the mature plant can vary based on the local condition (Atwater *et al*., 2016; Klein & Smith, 2021). McWhorter (1961) reported that the reproductive stage of johnsongrass started around a month after seed emergence, and the maximum rhizome growth was about 60 meters in 5 months. The flowering part of johnsongrass is a diffuse panicle, which is the primary feature in the image identification process. The flowering head is orange and purple at the mature stage.

Johnsongrass is a weedy relative of the cultivated sorghum (*Sorghum bicolor*), which compete for the same limiting resources, and the presence of johnsongrass will cause yield loss in sorghum or other crop fields (Hoffman & Buhler, 2002). Kansas and Texas are the top two states in cultivated sorghum production regarding planting acreage (NASS). As a weedy relative, johnsongrass is widespread in sorghum fields and along the roads around Nebraska, Kansas, and Texas (Ohadi *et al*., 2018; Werle Based on personal observation, johnsongrass is also widespread along the roads in other states such as California. Roadside populations can be mapped on a large scale with the integration of Google Street View Imagery and AI-based computer vision. In

addition to johnsongrass, this method can be applied to other weedy and invasive species. This project provided a workflow (Figure 1) to conduct an automated road survey of johnsongrass, examined the cost-effectiveness, and discussed the potential application of the johnsongrass population map.

1.6. Objectives

Since roadside weeds can spread to commercial crop fields, and also may cause traffic hazards, large-scale weed distribution maps are needed as the primary tool for designing weed management strategies. The traditional weed survey methods cannot create a large-scale species map because of the time and cost limitations. The AI-based object detection model can be an alternative method to car surveys. This study aimed to develop an automated AI-based virtual system to map weeds along roads in multiple states. Johnsongrass (*Sorghum halepense*) was used as the target species in the study.

The objectives of this project are:

- Create a training dataset by labeling GSV images that contain johnsongrass.
- Train a deep learning detection model of johnsongrass and map the distribution of the johnsongrass population along the roads.
- Evaluate and compare the cost-effectiveness of the AI-based survey against the traditional car survey.

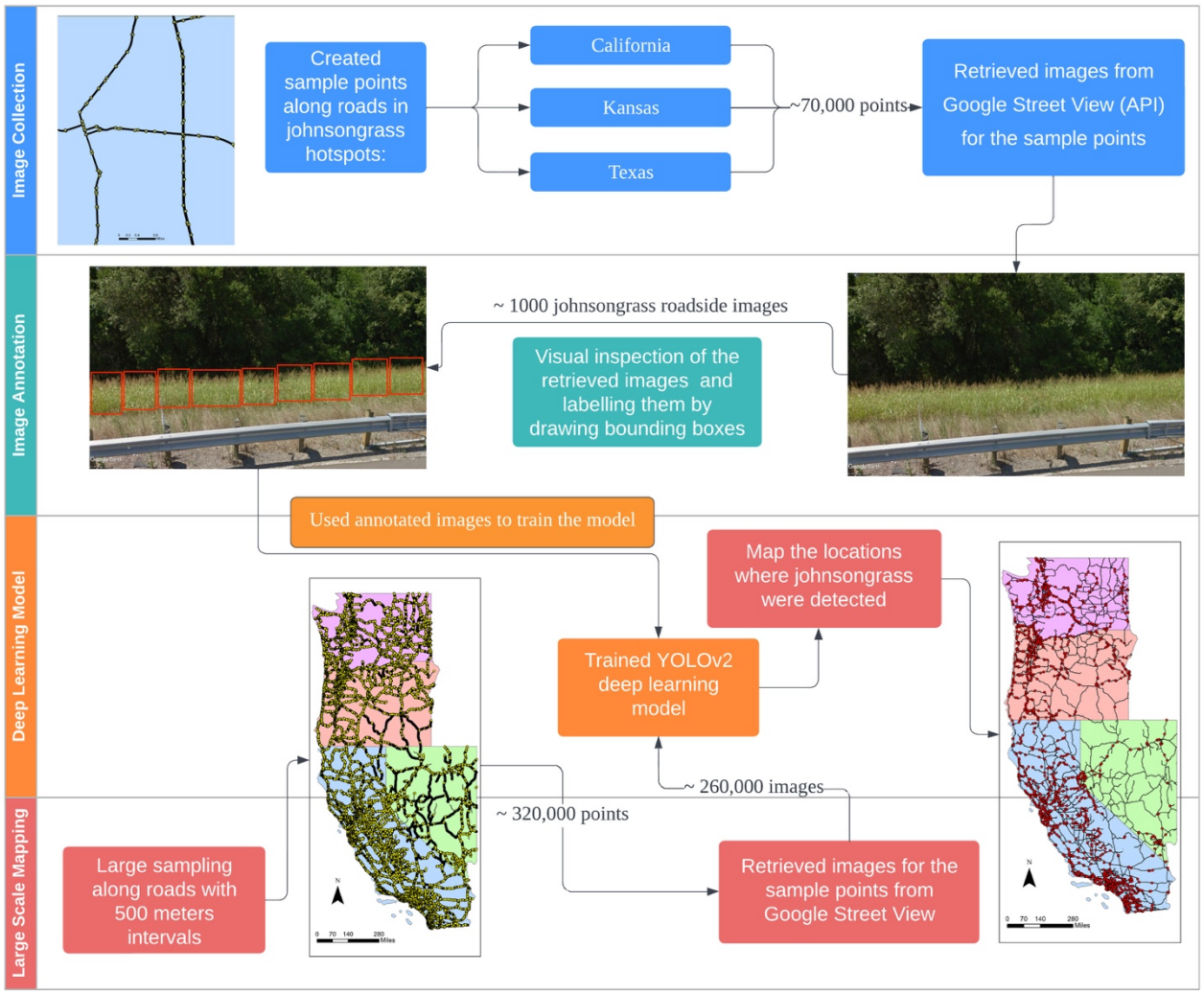

Figure 1. The overall workflow of the AI-based invasive species roadside survey. Images were retrieved from Google Street View and labeled in MATLAB 2021a. Map created with ArcGIS Pro Version 2.4.3 (Esri). Base map and road data retrieved from US Census Bureau: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.2021.html#list-tab-AGZ2ZC2D5ZN46GFI00

2. Materials and methods

2.1. Training Dataset

Since the model input was Google Street View (GSV) images, creating a training sample dataset with images from GSV could achieve better predicting performance. Training samples were collected by retrieving images from GSV in Texas, Kansas, and California because johnsongrass is widespread in those states, as noted in the previous section (Section 1.5). We retrieved 70,000 images from GSV API to obtain about 1000 images that include johnsongrass. Three parameters were specified when sending a request to GSV API, and the other parameters were set as default. The three parameters were latitude and longitude in the unit of decimal degree (e.g., 40.45, -80.00) and the heading. The heading indicates the camera angle based on the true compass heading. Google uses the value 0 and 360 to represent the North, 90 to represent the East, 180 to represent the South, and 270 to represent the West (Figure

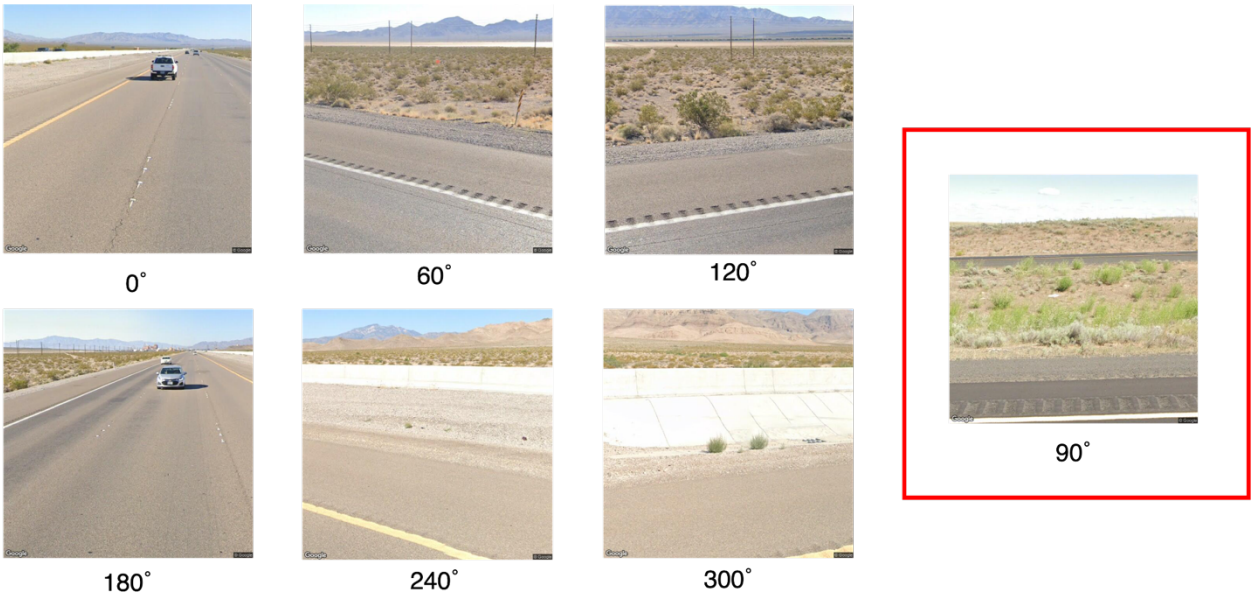

Figure 2. Pictures from Google Street View can be retrieved from different angles for a single location along the road heading north and south.

2). The picture perpendicular to (i.e.,90 degrees on a road heading north and south) the direction of the road will give the best view of the roadside. For roads not heading north and south, the angle was adjusted based on the direction the vehicle was traveling.

The images from GSV were grouped into a 1000-image folder, and each image was named by its location information (the value of latitude and longitude). MATLAB Image Processing Toolbox and Computer Vision Toolbox were used for the labeling process (The MathWorks, 2019). Inside the Computer Vision Toolbox, the Image Labeler was used to create annotated training data. This process aimed to identify the target species, johnsongrass, and draw a box around it. Johnsongrass identification was based on the red and orange flower heads and the tillers. When there was more than one individual johnsongrass in an image, multiple bounding boxes were drawn for each of them while ensuring that one bounding box only contained one individual, with no overlapping between boxes.

2.2. YOLO Object Detector

Many detection models based on neural networks have been introduced in the past ten years. Because the input was about 260,000 images (Figure 1), You Only Look Once (YOLO), a fast and accurate deep neural network, was applied in this project. YOLO is a real-time image detection model, and it can handle images at 45 frames per second in the base model and 155 frames per second in the smaller version of the network called Fast YOLO (Redmon *et al*., 2016). YOLO is a popular detection network, and it has been applied to many weedy and invasive species studies (Dang *et al*., 2022; Parico & Ahamed, 2020; Wang *et al*., 2022). In this project, the updated version of

YOLO was adopted, called YOLOv2, which improved performance and could process up to 91 frames per second when the input size is 288 × 288 (Redmon & Farhadi, 2017). In comparison, Faster R-CNN, another well-known deep learning model, can only process five frames per second in a deep detection model and 17 frames per second in a shallow model (Ren *et al*., 2017). Faster R-CNN applied the idea of the Region Proposal Network, which means dissecting the image into several rectangular proposals and generating many small boxes called anchor boxes with different widths and heights at the center of each rectangle (Ren *et al*., 2017). Instead of applying Region Proposal Network, YOLOv2 dissected the image into different grids and predicted the location of the bounding boxes by applying much fewer anchor boxes than Faster R-CNN (Redmon & Farhadi, 2017). The number of anchor boxes is the main factor for a faster computational speed.

ResNet50 (50 layers) was adopted as the feature extraction network and paired with the YOLOv2 detector (He *et al*., 2016). Recent neural networks often develop layers from 100 to 1000 or even more. He *et al*. (2016) denoted an input x and the output f(x) in a single neural network layer, and ResNet would take f(x) - x, the residual of the output, as the input for the following neural network layer. Minimizing the input in each layer will significantly improve the performance of a deep neural network. Zheng *et al*. (2019) denoted that ResNet50 has the best performance on the dataset CropDeep among seven different classification models (VGG 16, VGG19, SqueezeNet, Inception V4, DenseNet121, ResNet18, and ResNet50), and the average accuracy was 99.81. Abdulsalam & Aouf (2020) tested the integration of YOLOv2 and ResNet50, and the accuracy of detection and classification on Bluegrass, Chenopodium, and Cirsium was

99.5%, but the accuracy on Sedges was about 84%. These two studies suggested that ResNet50 might be the best model to classify weedy species in an image. Integrating YOLOv2 and ResNet50 could achieve large-scale weed detection along the road in a short time with high accuracy.

2.3. Model Training

The training dataset included 911 images with johnsongrass, where 75% of the images were used as the training dataset and 25% as the validation dataset. The input image size was 224 × 224 × 3, which indicated the height and width of the image and three color bands, respectively. Different numbers of anchors were applied and compared to determine the anchor number that gives the highest Intersection over Union (IoU) value. IoU value is the area of overlap between predicted boxes and the labeled bounding box (ground truth). The number of anchor boxes was set at 12 in the training model as it provided the best IoU value. We adopted ResNet50 as the feature extraction network, and the activation function was "activiation_40-relu" in MATLAB Deep Learning Toolbox (The MathWorks, 2019).

The size of the training samples is important in deep learning models, and data augmentation can significantly boost the training sample size. The data from Dang *et al*. (2022) proved that augmentation could improve the accuracy of the YOLO model in terms of metrics like recall and mean Average Precision (mAP). Data augmentation creates extra training samples by changing the image orientation, rotation, contrast, saturation, and hue. For example, as shown in Figure 3, four versions of the same image were created by randomly changing the contrast, hue, saturation, and brightness

in a set range and flipping the original image.

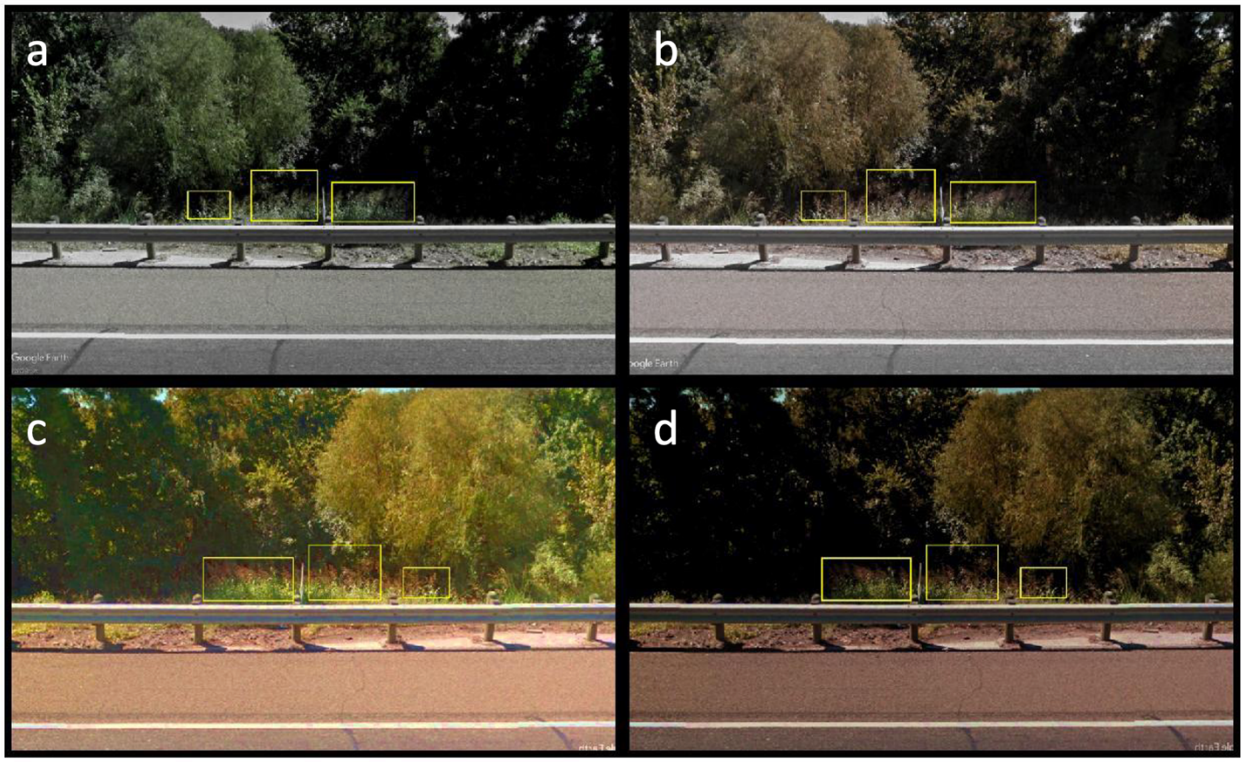

Figure 3. Examples of the augmented training datasets. (a) and (b) randomly changed the contrast, hue, saturation, and brightness of the original image. (c) and (d) randomly changed the contrast, hue, saturation, and brightness and flipped the original image.

To train the model, we applied *Adam*, an optimization algorithm for the stochastic objective function, best known for its high computational efficiency and ease of use (Kingma & Ba, 2014). The model had a mini-batch size of 16 with a 0.001 initial learning rate and a maximum number of epochs of 30. Mini-batch size is the number of images we divide into a group to send to the learning model, and the initial learning rate is the step size for each iteration to reach the bottom of a loss function, while epoch indicates the number of cycles the learning model uses the whole set of training samples (Géron, 2019; Kingma & Ba, 2014).

2.4. Model Testing and Performance Evaluation

The trained detection model was tested by applying the testing dataset, which included 2,040 GSV images. A human identification test was conducted on the testing dataset as well. The results from the detection model and human observers were compared. A confusion matrix was used to evaluate the model performance. The four values in the cells are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

The model was evaluated based on overall accuracy, precision, recall, and false positive rate (FPR), the standard evaluation metrics in model performance (Dang *et al*., 2022; Parico & Ahamed, 2020; Wang *et al*., 2022). Precision is the percentage of true positives among all the detections, and recall is the percentage of true positives among all the ground-truth objects (Padilla *et al*., 2020). Accuracy is an overall measurement of the percentage of true positives and true negatives in the entire dataset (Story & Congalton, 1986). The formulas to calculate precision, recall, FPR, and accuracy are listed below (Padilla *et al*., 2020; Story & Congalton, 1986):

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (1)$$

$$Recall = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})} \quad (2)$$

$$False\ Positive\ Rate\ (FPR) = \frac{False\ Positives}{False\ Positives + True\ Negatives} \quad (3)$$

$$Overall\ Accuracy = \frac{(True\ Positives + True\ Negatives)}{(True\ Positives + False\ Positives + True\ Negatives + False\ Negatives)} \quad (4)$$

2.5. Weed Map

After the training and validation processes, the final step was to create a johnsongrass map showing the populations along the roads in California, Oregon, Washington, and Nevada. First, the base road map (Figure 4a) was extracted from the US Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (US Census Bureau, 2021). In Figure 4a, the black lines are primary and secondary roads, and the total length of these roads across the study area is 135,000 km.

A total of 320,000 sampling points along the roads were selected on the map (Figure 4b), and each sampling point was 500 m apart along the roads. Every point's geographical coordinate (latitude and longitude) was retrieved and saved. The location information of every single sample point was used to retrieve 269,489 images instead of 320,000 through GSV API because GSV images are unavailable in some locations. The trained model detected the input images and outputted a bounding box with the probability score around the target if the target is present in the image. Finally, a map of the target species was created with ArcGIS Pro Version 2.4.3 (Esri) using the coordination data from the sampling points that had johnsongrass.

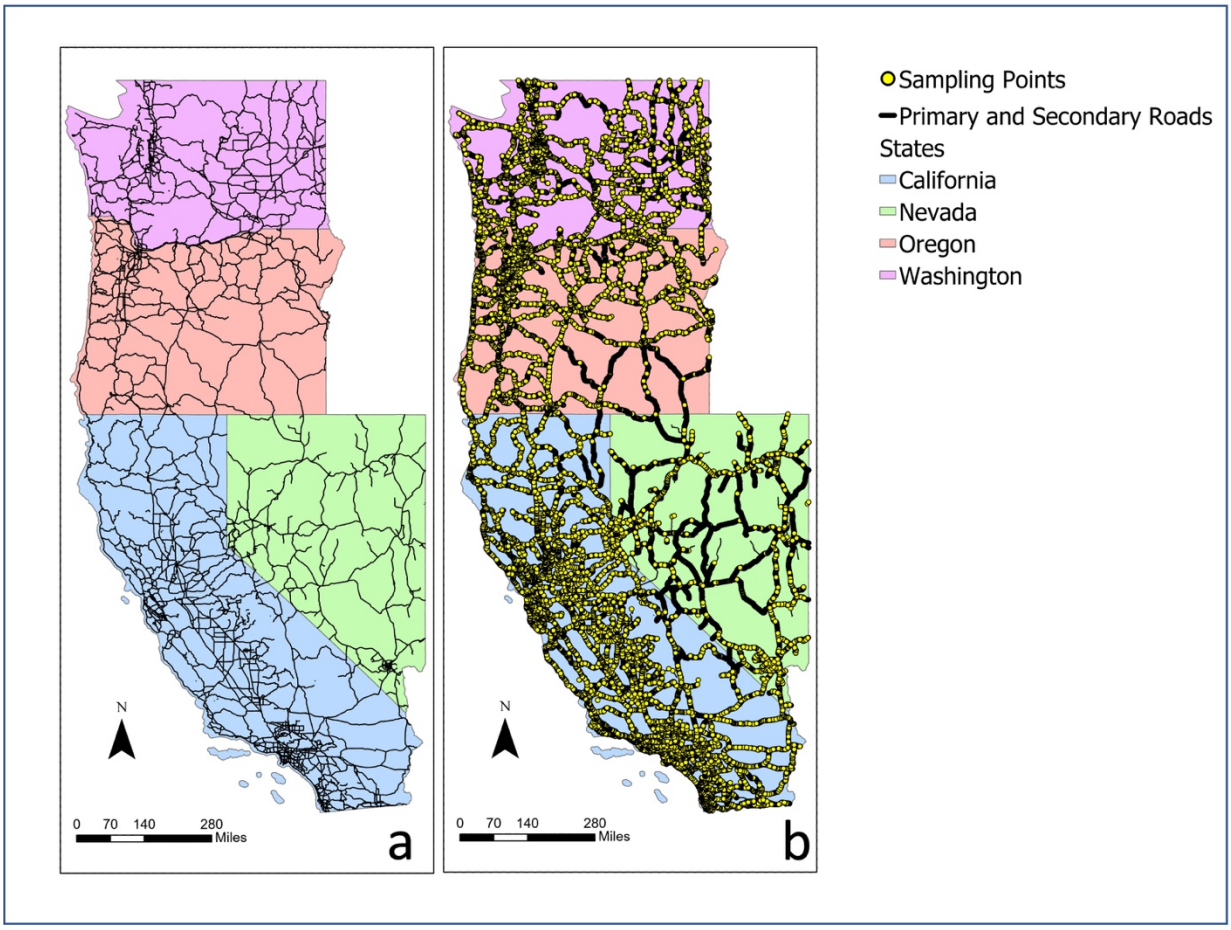

Figure 4. (a) The primary and secondary roads in California, Nevada, Oregon, and Washington in the United States. (b) Sampling points ,500 meters apart, used for image retrieval from GSV. Map created from ArcGIS Pro Version 2.4.3 (Esri). Base map and road data retrieved from US Census Bureau: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.2021.html#list-tab-AGZ2ZC2D5ZN46GFI00

3. Results and Discussion

3.1. Training results

After training the model with the augmented training dataset, the original images were labeled with three items. For example, Figure 5 shows the red bounding box as human-labeled data (ground truth). The yellow box indicates the detections of johnsongrass with bounding boxes and the confidence scores. In this project, we set the threshold confidence score as 0.6, so any detection with a score greater than or equal to 0.6 will be identified as johnsongrass. For example, in Figure 5c, the scores of the two detected targets are 0.656 and 0.728, and the model identified them as johnsongrass, which is a true positive in the context of the confusion matrix.

Dang *et al*. (2022) and Parico & Ahamed (2020) applied YOLO detection models with 0.5 thresholds on UAV or aerial image weed detections. The threshold value was

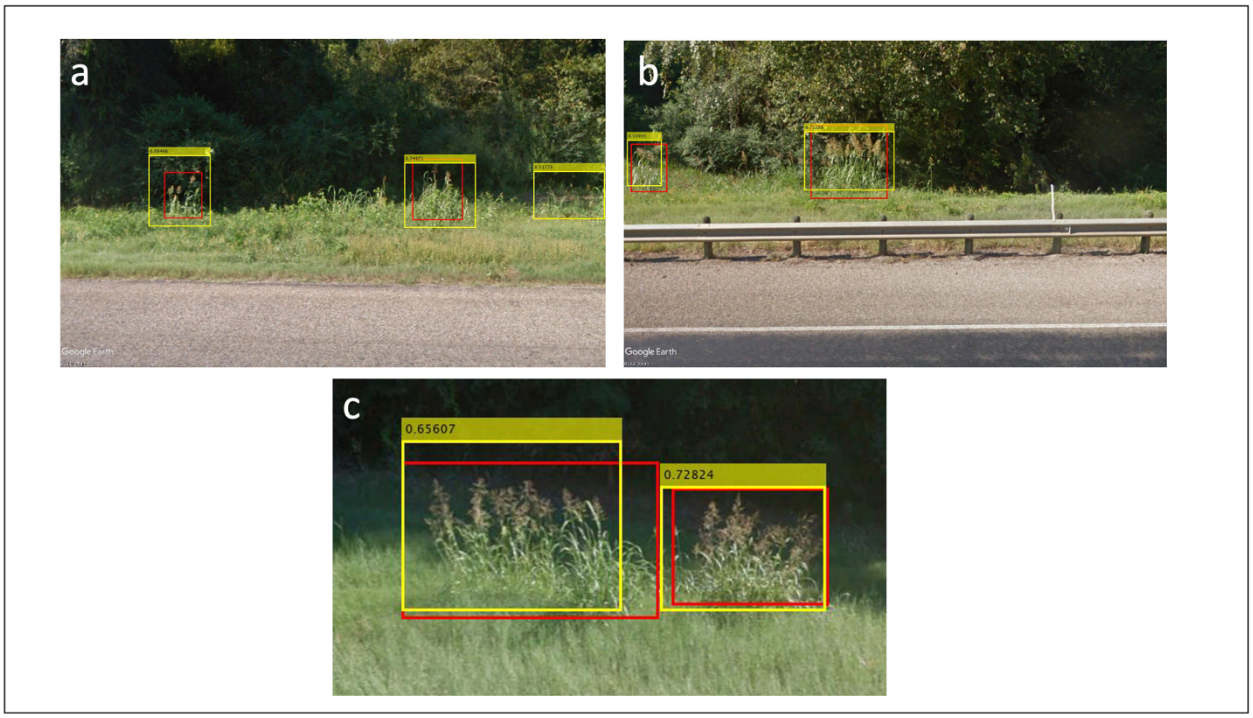

Figure 5. Examples of detections of johnsongrass (Sorghum halepense) on Google Street View images by YOLO v2 model in the validation process. Red boxes are human-labelled in original training dataset. Yellow boxes and the scores on top are the output of the detection model.

set to around 0.5 to account for human errors in the labeling process (Parico & Ahamed, 2020). In our study, the threshold value was set to 0.6 to reduce the false positive errors.

3.2. Model performance

The performance of the YOLOv2 model in detecting johnsongrass in GSV images was tested using a total of 2,040 test images. Based on the threshold value of 0.6 for presence (i.e., johnsongrass is present in the image), the confusion matrix shown in Table 1 was created. The YOLOv2 model achieved a recall value of 85% in the GSV testing dataset. Dang *et al*. (2022) reported similar recall values in their study which the average recall value of the YOLOv3 detection model on 12 different weed species was 87.93%. However, there is still about a 15% chance that johnsongrass could be undetected by our model when johnsongrass is present in the image. In the testing dataset, there were 153 images were classified as FN in the confusion matrix. FN in this project is the image that contained johnsongrass and was identified and labeled by human observers, but the model was not able to output the same result. Individuals that were at pre-flowering stage were not considered as FN since the training dataset only contained mature johnsongrass.

The precision for the YOLOv2 model was 0.74, which is lower than the recall (Table 2). Both the precision and FPR include false positive detection in the calculation. FPR (0.3) implies that the model could wrongly detect other plant species as johnsongrass, with a 30% chance. Among the group of incorrect detection (FP + FN), FP had twice the number of images compared to FN in the test dataset (Table 1). Since

most studies in weed detection were conducted in the crop field, and their models were applied to distinguish weeds from the crop, the precision values were high and were about 85% to 95% (Dang *et al*., 2022; Wang *et al*., 2022). Yan & Ryu (2021) applied a CNN model on GSV images to detect roadside crop type, and the results denoted that most crops had detection precision above 90%, but only rice (*Oryza sativa*) had a 76% precision. The study also reported that the misclassification of rice was more frequent in low-resolution images or if the object was far away from the camera (Yan & Ryu, 2021). In our research, the quality of images and the distance between the target object and the camera might contribute to the high FP value. Future work will focus on decreasing the rate of FP in the image where johnsongrass is absent. More CNN models will be tested on the johnsongrass training database to compare the precision and accuracy of different models. More roads with high-resolution images from Google can help improve this survey method's accuracy.

The overall accuracy of the YOLOv2 model was 77.5% for detecting johnsongrass in the GSV images. This index provided an overall evaluation based on total correct detection and the total number of test images. Ringland *et al*. (2019) and Yan & Ryu (2021) both conducted image detection models on GSV images to survey different types of crop production along the roads but with different CNN networks from our model. The accuracy of detecting general crops like alfalfa (*Medicago sativa*), almond (*Prunus dulcis*), corn (*Zea Mays*), and rice in the GSV images could reach 92% (Yan & Ryu, 2021). An explanation of high accuracy on crops is that major crops always have unique morphology or patterns because of domestication, row spacing, and field layout that might help to increase performance in computer vision. For roadside weedy

species like johnsongrass, morphological variations under different environmental conditions were reported in many studies, and the variation could lead to low precision and overall accuracy (Atwater *et al*., 2016; Klein & Smith, 2021).

There were several challenges in the labeling process, and they can explain most of the incorrect detections. In some annotated training images, the target species were partially occluded by other objects, including other invasive species, traffic signs, and fences. In this case, we could only label either the flowering part or the basal part of johnsongrass. In this project, and for johnsongrass specifically, the panicle part of the plant would be labeled in most cases since we could not differentiate the basal part of johnsongrass from other grass species. The growing stages of the target species were another challenge in the labeling process. The juvenile stage of johnsongrass has no panicles and looks similar to many other grass species. Only the mature johnsongrasses were included in the training dataset, so the model was unlikely to detect individuals at their early vegetative stage.

Table 1. Confusion matrix of YOLOv2 model on the testing dataset (n = 2,040). TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative

		Predicted	
	Confusion Matrix (n = 2,040)	Johnsongrass Present	Johnsongrass Absent
Actual	Johnsongrass Present	867 (TP)	153 (FN)
	Johnsongrass Absent	306 (FP)	714 (TN)

Table 2. Model Performance Metrics of YOLOv2 model on the testing dataset (n = 2,040). FPR: False Positive Rate.

Total Number	Recall	Precision	FPR	Overall Accuracy
2,040	85%	73.9%	30%	77.5%

3.3. Johnsongrass mapping

The trained model was applied to 269,489 images collected from Google Street View. In Figure 6, the red points denote the potential location of johnsongrass predicted by the model. The model identified a total of 2,031 images as having johnsongrass. The predicted distribution of johnsongrass suggested that johnsongrass is less widespread in Nevada than in the other states in this study. The location shown on this map is only the prediction. The johnsongrass individual might not be found in that location depending on the growing season since most images were taken 2 to 3 years ago. Deus *et al*. (2016) conducted a Google Street View study that surveyed *E. globulus* (Tasmanian blue gum), and their results mentioned that environmental stresses (e.g., frost) could lead to variability in species abundance in a short period, from one to two years.

Recent studies and our results suggested that integrating GSV and a deep learning image detection model can map species on a much larger scale. Yan & Ryu (2021) integrated GSV and other deep learning algorithms and produced cropping system maps of Central Valley in California and the state of Illinois. Another roadside crop survey in Thailand covered 572 km of road and examined about 57,000 panoramas (Ringland *et al*., 2019). Our study covered more areas (California, Oregon, Nevada, and Washington), longer roads (135,000 km), and more panoramas (269,489) than studies used a similar road survey method (Ringland *et al*., 2019; Yan & Ryu, 2021). Future research will focus on the survey in other states in the US, and our goal is to survey all the roads in the US.

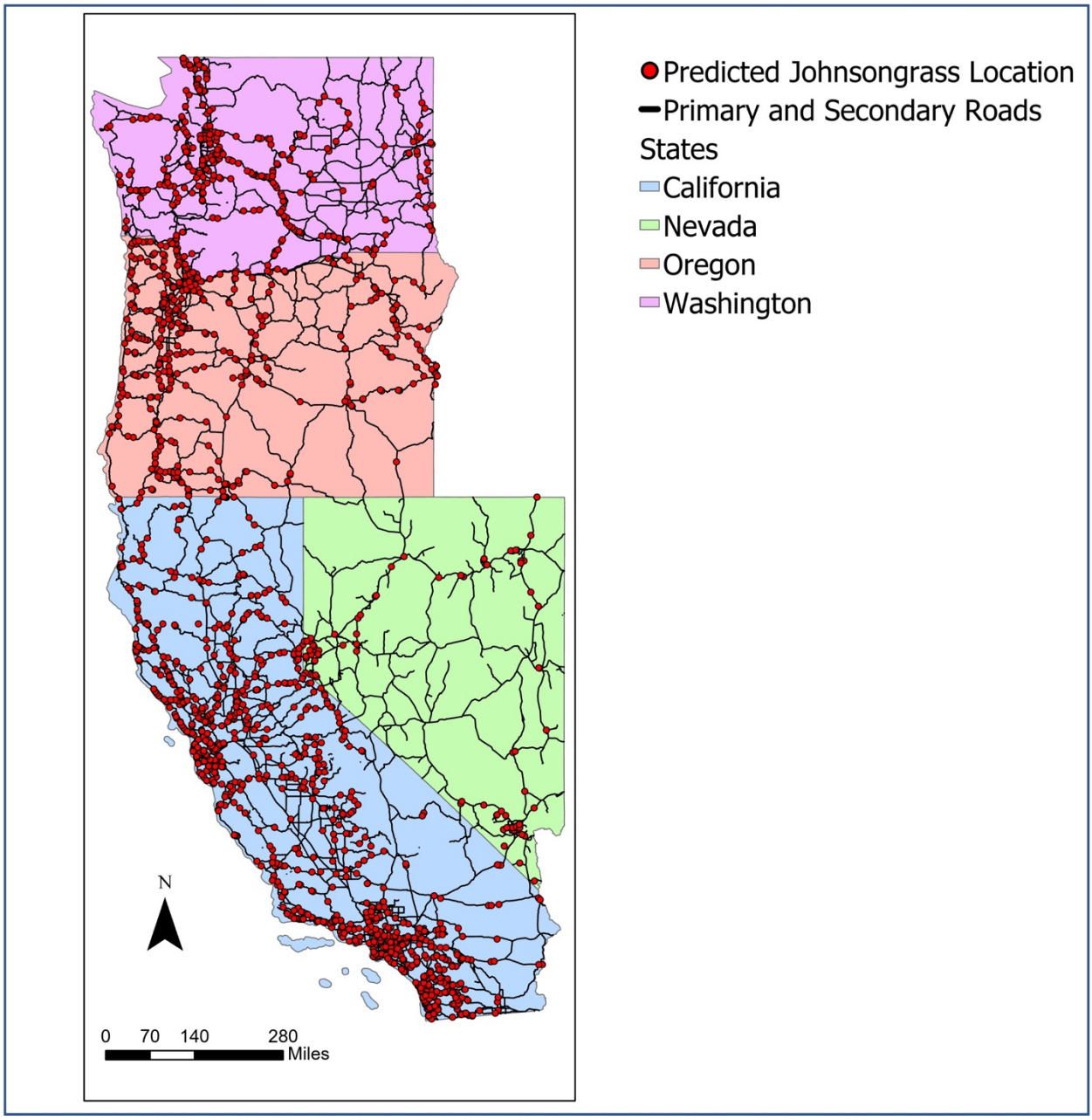

Figure 6. Johnsongrass distribution map (red dots) produced via deep learning model. Map created from ArcGIS Pro Version 2.4.3 (Esri). Base map and road data retrieved from US Census Bureau: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.2021.html#list-tab-AGZ2ZC2D5ZN46GFI00

YOLO has been tested in many studies to detect multiple plant species in a single image (Dang et al., 2022; Ringland et al., 2019). Johnsongrass was the only detection target in this study, but other invasive species can be mapped by using our methods. A larger-scale species distribution map can be combined with environmental factors or land use to determine the conditions suitable for spreading the species. An example would be the habitat suitability model, which predicts how well species thrive and spread in a location given environmental conditions (Hirzel *et al*., 2006). According to Crall *et al*. (2013), even though the habitat suitability model is a key tool for invasive species risk management, the model requires location data on a large spatial scale. Our method can provide a more prominent presence/absence dataset than the traditional local dataset. Habitat suitability models based on a larger scale can yield a more robust conclusion. As noted, the sampling created by our method is a biased sampling of the environment under which the species can thrive as we only search for species along the roadside habitats.

AI-based surveys can provide accurate location data to build and test invasive species dispersal models. The AI-based mapping approach can only detect roadside weedy species. A dispersal model can be applied based on the johnsongrass location map. A typical dispersal model requires two primary parameters, reproduction rate and spread distance, and then for parameterization and calibration, a multiple-time-step map is required (Adams *et al*., 2015). The johnsongrass map was created based on images from a different date. Even though most of the images were taken in recent years, like 2020, a small portion was taken 8 or 10 years ago. Based on the API method to retrieve images from GSV, we cannot choose the date of the image from the exact location, but

Google will always return the latest image from the same location. As a result, we are unable to create multiple-time-step maps based on the AI-based method. Ringland *et al.* (2019) reported a similar limitation on Google API image acquisition. The multiple-time-step may become available in the future by Google. The time and location of the GSV update are listed on their website (https://www.google.com/intl/en/streetview/).

3.4. The cost-effective approach

Expenses and estimated time for the car survey, the human-based GSV survey, and the AI-based GSV survey were calculated (Table 3). Expenses for the car survey were calculated based on the same scale (135,000 km) as the AI-based survey. The cost breakdown of the car survey was calculated based on regular domestic travel daily expenses. Vehicle rentals and gas estimation are US$ 9,408 and US$ 7,350, respectively, and the accommodation accounts for a more significant portion of the costs, i.e., US$ 25,200 (Table 3). The total travel time of a car survey per person requires 180 days, estimated based on a daily 750 km drive. Dues *et al*. (2016) conducted a 38-day car survey of 15,000 km of roads in Portugal, and a standard car survey would drive much less than 750 km per day. For the human- and AI-based GSV survey, US$1,890 is required for image purchase from Google, which is US$7 per 1000 images (Table 3). Labor costs for all three types of surveys were calculated based on the minimum hourly wage in California and the total time spent on each method. In terms of cost, the AI-based GSV survey is 50% less than the human-based GSV survey, and the AI-based GSV survey is only 5.6% of the total cost of the car survey.

Compared to car surveys, GSV-based surveys did not require outdoor work and driving, so GSV-based detection can minimize potential worker risks (e.g., car accidents and health-related risks). Studies by Deus *et al*. (2016) and Kotowska *et al*. (2021) reported that the results produced by GSV resemble that produced by the field survey. Compared to the GSV survey by a human observer, the AI-based GSV survey had spent shorter time. Once the detection model is trained, the machine can work 24 hours per day, but on average, a human can process about 6000 ~ 8000 images per day based on our labeling experiences. A more detailed GSV survey might take more time. For example, a human-based GSV survey took 35 hours to examine 2,350 panoramas in Sicily, Italy, to assess invasive species abundance along the roads (Barone *et al*., 2021).

As noted, once the training dataset is created, the labor cost of the AI-based GSV surveys is fixed and will not increase as the number of images increases. However, the relation between labor cost and image number in the human-based surveys is linear. As the sampling scales increase, AI-based surveys will outcompete human-based surveys, and the comparison between these two surveys in Table 3 is underestimated. Then in terms of errors, the detection errors by the AI-based model could be consistent, quantified, and improved, while errors in car surveys and human-based GSV surveys are unable to quantify and inconsistent. The AI-based method is much more cost-effective than the car survey and human-based GSV surveys for large-scale species surveys.

Because of the strong effects of climate changes, species' habitats and favorable environmental conditions can change rapidly, and the changing climate may accelerate

invasive species spread (Chai *et al*., 2016). Based on the AI-based species mapping method and the integration with other ecological models, we can predict the spread of invasive species and monitor the population closely. The prediction can then allow us to allocate our resources better to control the spread of invasive species.

Table 3. Cost comparison between car survey, human-vision GSV survey and computer-vision AI-based survey for a 135,000 km species road survey.

	Expenses	**Amount (US dollars)**	**Time Spent (days)**
Car Survey	Vehicle Rentals	9,408	N/A
	Fuel	7,350	N/A
	Meals and Hotels	25,200	N/A
	Labor (Hourly Pay)	21,600	N/A
	Total	**63,558**	**180**
GSV Virtual Survey	Image Cost (269,489 images)	1,890	N/A
	Labor (Hourly Pay)	3,960	N/A
	Total	**5,850**	**33**
AI-based Survey	Image Cost (269,489 images)	1,890	N/A
	Labor (Hourly Pay)	1,680	N/A
	Total	**3,570**	**14**

4. Conclusion

Our study developed a novel weed survey method integrating GSV and a deep learning model. The overall accuracy of johnsongrass detection in GSV panoramas could achieve 77.5% with an 85% recall and 73.9% precision. The recall could be improved by adding more training samples of johnsongrass in the early vegetative stage. However, young johnsongrass is difficult to identify by human as well, and the labor of labeling will increase. The precision could be improved by utilizing different CNN image detection models. Our trained model detected 2,031 images with johnsongrass presence out of 269,489 GSV images in California, Oregon, Washington, and Nevada. The locations of those 2,031 images were used to create a distribution map of johnsongrass along the major roads in these four states. We explained and gave examples of possible potential applications and further data analysis based on the johnsongrass map we created. We compared the cost of the car survey, human-based GSV survey, and AI-based GSV survey, and the result demonstrated that the AI-based GSV survey could spend much less time and money on larger-scale roadside species mapping. However, our method cannot retrieve images on a specific date for a single location as the current Google API always returns the latest images to us from their database. Our methods can be improved as the GSV database increase the size and provide more options for images from a single location but at different time. Overall, our work presented that an AI-based GSV survey can be cost-effective on roadside invasive species surveys. Future work will focus on expanding the scale of johnsongrass mapping, and our model will be applied to other invasive species to create more large-scale distribution maps.

Literature Cited

Abdulsalam, M., & Aouf, N. (2020). Deep weed detector/classifier network for precision agriculture. *2020 28th Mediterranean Conference on Control and Automation (MED)*. http://dx.doi.org/10.1109/med48518.2020.9183325

Abella, S. R., Spencer, J. E., Hoines, J., & Nazarchyk, C. (2008). Assessing an exotic plant surveying program in the Mojave Desert, Clark County, Nevada, USA. *Environmental Monitoring and Assessment*, *151*(1–4), 221–230. https://doi.org/10.1007/s10661-008-0263-0

Adams, V. M., Petty, A. M., Douglas, M. M., Buckley, Y. M., Ferdinands, K. B., Okazaki, T., Ko, D. W., & Setterfield, S. A. (2015). Distribution, demography and dispersal model of spatial spread of invasive plant populations with limited data. *Methods in Ecology and Evolution*, *6*(7), 782–794. https://doi.org/10.1111/2041-210x.12392

Atwater, D. Z., Sezen, U. U., Goff, V., Kong, W., Paterson, A. H., & Barney, J. N. (2015). Reconstructing changes in the genotype, phenotype, and climatic niche of an introduced species. *Ecography*, *39*(9), 894–903. https://doi.org/10.1111/ecog.02031

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google street view: Capturing the world at street level. *Computer*, *43*(6), 32–38. https://doi.org/10.1109/mc.2010.170

Ansong, M., & Pickering, C. (2013). Are weeds hitchhiking a ride on your car? A systematic review of seed dispersal on cars. *PLoS ONE*, *8*(11), e80275. https://doi.org/10.1371/journal.pone.0080275

Bagavathiannan, M. V., & Norsworthy, J. K. (2016). Multiple-Herbicide resistance is widespread in roadside *Palmer Amaranth* populations. *PLOS ONE*, *11*(4), e0148748. https://doi.org/10.1371/journal.pone.0148748

Baker, H. G. (1974). The evolution of weeds. *Annual review of ecology and systematics*, 1-24.

Barone, G., Domina, G., & Di Gristina, E. (2021). Comparison of different methods to assess the distribution of alien plants along the road network and use of Google Street View panoramas interpretation in Sicily (Italy) as a case study. *Biodiversity Data Journal*, *9*.

Bhatti, A. G., Endrizzi, J. E., & Reeves, R. G. (1960). Origin of Johnson grass. *Journal of Heredity*, *51*(3), 107–110. https://doi.org/10.1093/oxfordjournals.jhered.a106962

Bohemen, H. D. V., & Janssen van de Laak, W. H. (2003). The influence of road infrastructure and traffic on soil, water, and air quality. *Environmental Management, 31*(1), 50–68. https://doi.org/10.1007/s00267-002-2802-8

Brooks, M. L., D'Antonio, C. M., Richardson, D. M., Grace, J. B., Keeley, J. E., DiTOMASO, J. M., Hobbs, R. J., Pellant, M., & Pyke, D. (2004). Effects of invasive alien plants on fire regimes. *BioScience*, *54*(7), 677. https://doi.org/10.1641/0006-3568(2004)054[0677:eoiapo]2.0.co;2

California Department of Transportation. (n.d.). Roadside vegetation control. Caltrans. Retrieved September 12, 2022, from https://dot.ca.gov/caltrans-near-me/district-4/d4-projects/roadside-vegetation-control

California Invasive Plant Council. (n.d.). *About invasive plants*. Retrieved June 4, 2022,

from https://www.cal-ipc.org/plants/impact/

California Invasive Plant Council. (2017). *The cost of invasive plants on California*. Retrieved September 12, 2022, from https://www.cal-ipc.org/solutions/research/cost/

Catry, F. X., Moreira, F., Deus, E., Silva, J. S., & Águas, A. (2015). Assessing the extent and the environmental drivers of *Eucalyptus globulus* wildling establishment in Portugal: Results from a countrywide survey. *Biological Invasions*, *17*(11), 3163–3181. https://doi.org/10.1007/s10530-015-0943-y

Chai, S.-L., Zhang, J., Nixon, A., & Nielsen, S. (2016). Using risk assessment and habitat suitability models to prioritise invasive species for management in a changing climate. *PLOS ONE*, *11*(10), e0165292. https://doi.org/10.1371/journal.pone.0165292

Crall, A. W., Jarnevich, C. S., Panke, B., Young, N., Renz, M., & Morisette, J. (2013). Using habitat suitability models to target invasive plant species surveys. *Ecological Applications*, *23*(1), 60–72. https://doi.org/10.1890/12-0465.1

Crystal-Ornelas, R., Hudgins, E. J., Cuthbert, R. N., Haubrock, P. J., Fantle-Lepczyk, J., Angulo, E., Kramer, A. M., Ballesteros-Mejia, L., Leroy, B., Leung, B., López-López, E., Diagne, C., & Courchamp, F. (2021). Economic costs of biological invasions within North America. *NeoBiota, 67*, 485–510. https://doi.org/10.3897/neobiota.67.58038

Dang, F., Chen, D., Lu, Y., Li, Z., & Zheng, Y. (2022). DeepCottonWeeds (DCW): A Novel Benchmark of YOLO Object Detectors for Weed Detection in Cotton Production Systems. In *2022 ASABE Annual International Meeting* (p. 1).

American Society of Agricultural and Biological Engineers.

Das, K., & Duarah, P. (2013). Invasive alien plant species in the roadside areas of Jorhat, Assam: Their harmful effects and beneficial uses. *International Journal of Engineering Research and Applications, 3*(5), 353-358.

Davidson, A. M., Jennions, M., & Nicotra, A. B. (2011). Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecology Letters, 14*(4), 419–431. https://doi.org/10.1111/j.1461-0248.2011.01596.x

Deus, E., Silva, J. S., Catry, F. X., Rocha, M., & Moreira, F. (2016). Google Street View as an alternative method to car surveys in large-scale vegetation assessments. *Environmental Monitoring and Assessment*, *188*(10). https://doi.org/10.1007/s10661-016-5555-1

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, *151*, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Dyrmann, M., Jørgensen, R. N., & Midtiby, H. S. (2017). RoboWeedSupport - Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences*, *8*(2), 842–847. https://doi.org/10.1017/s2040470017000206

Federal Highway Administration (FHWA). (n.d.). *Office of Highway Policy Information - Policy*. Federal Highway Administration. Retrieved June 9, 2022, from https://www.fhwa.dot.gov/policyinformation/statistics/2008/hm60.cfm

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a

mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/bf00344251

Gelbard, J. L., & Belnap, J. (2003). Roads as conduits for exotic plant invasions in a semiarid landscape. *Conservation Biology*, *17*(2), 420–432. https://doi.org/10.1046/j.1523-1739.2003.01408.x

Géron, A. (2019). *Hands-On machine learning with Scikit-Learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems*. "O'Reilly Media, Inc."

Girshick, R. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*. http://dx.doi.org/10.1109/iccv.2015.169

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. http://dx.doi.org/10.1109/cvpr.2014.81

Hakkim, V., Joseph, E., Gokul, A., & Mufeedha, K. (2016). Precision farming: The future of Indian agriculture. *Journal of Applied Biology and Biotechnology*, *4*(06), 068–072. https://doi.org/10.7324/jabb.2016.40609

Hardion, L., Leriche, A., Schwoertzig, E., & Millon, A. (2016). Species distribution 2.0: An accurate time- and cost-effective method of prospection using street view imagery. *PLOS ONE*, *11*(1), e0146899. https://doi.org/10.1371/journal.pone.0146899

He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Herold, J., Lowe, Z., & Dukes, J. (2014). *Integrated vegetation management (IVM) for INDOT roadsides* (No. FHWA/IN/JTRP-2013/08). Purdue University. Joint Transportation Research Program.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological modelling*, *199*(2), 142-152.

Hoffman, M. L., & Buhler, D. D. (2002). Utilizing Sorghum as a functional model of crop–weed competition. I. Establishing a competitive hierarchy. *Weed Science*, *50*(4), 466-472.

Hyman, W. A., & Vary, D. (1999). *Best management practices for environmental issues related to highway and street maintenance*. Transportation Research Board.

Jodoin, Y., Lavoie, C., Villeneuve, P., Theriault, M., Beaulieu, J., & Belzile, F. (2008). Highways as corridors and habitats for the invasive common reed *Phragmites australis* in Quebec, Canada. *Journal of Applied Ecology*, *45*(2), 459–466. https://doi.org/10.1111/j.1365-2664.2007.01362.x

Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, *14*(2), 401–413. https://doi.org/10.1890/02-5364

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

Klein, P., & Smith, C. M. (2021). Invasive Johnsongrass, a threat to native grasslands and agriculture. *Biologia*, *76*(2), 413–420. https://doi.org/10.2478/s11756-020-00625-5

Kotowska, D., Pärt, T., & Żmihorski, M. (2021). Evaluating Google Street View for tracking invasive alien plants along roads. *Ecological Indicators*, *121*, 107020. https://doi.org/10.1016/j.ecolind.2020.107020

Loebach, C. A., & Anderson, R. C. (2018). Measuring short distance dispersal of *Alliaria petiolata* and determining potential long distance dispersal mechanisms. *PeerJ*, *6*, e4477. https://doi.org/10.7717/peerj.4477

McWhorter, C. G. (1961). Morphology and development of johnsongrass plants from seeds and rhizomes. *Weeds*, *9*(4), 558. https://doi.org/10.2307/4040804

Mills, S. D., Mamo, M., Ruis, S. J., Blanco-Canqui, H., Schacht, W. H., Awada, T., Li, X., & Sutton, P. (2021). Soil properties limiting vegetation establishment along roadsides. *Journal of Environmental Quality*, *50*(1), 110–121. https://doi.org/10.1002/jeq2.20184

Milton, S. J., & Dean, W. R. J. (1998). Alien plant assemblages near roads in arid and semi-arid South Africa. *Diversity and Distributions*, *4*(4), 175-187.

Mortensen, D. A., Rauschert, E. S. J., Nord, A. N., & Jones, B. P. (2009). Forest roads facilitate the spread of invasive plants. *Invasive Plant Science and Management*, *2*(3), 191–199. https://doi.org/10.1614/ipsm-08-125.1

Nathan, R. (2006). Long-Distance dispersal of plants. *Science*, *313*(5788), 786–788. https://doi.org/10.1126/science.1124975

Neher, D. A., Asmussen, D., & Lovell, S. T. (2013). Roads in northern hardwood forests affect adjacent plant communities and soil chemistry in proportion to the maintained roadside area. *Science of The Total Environment, 449*, 320–327. https://doi.org/10.1016/j.scitotenv.2013.01.062

Ohadi, S., Littlejohn, M., Mesgaran, M., Rooney, W., & Bagavathiannan, M. (2018). Surveying the spatial distribution of feral sorghum (*Sorghum bicolor L.*) and its sympatry with johnsongrass (*S. halepense*) in South Texas. *PLOS ONE*, *13*(4), e0195511. https://doi.org/10.1371/journal.pone.0195511

Othman Ghafoor, A. (2010). ALLELOPATHIC EFFECTS OF JOHNSON GRASS (Sorghum halepense L.) ON GERMINATION AND SEEDLING GROWTH OF SOME AND LEGUMES CROPS CRREALS. *Mesopotamia Journal of Agriculture*, *38*(2), 174–180. https://doi.org/10.33899/magrj.2010.27845

Parendes, L. A., & Jones, J. A. (2000). Role of light availability and dispersal in exotic plant invasion along roads and streams in the H. J. Andrews Experimental Forest, Oregon. *Conservation Biology*, *14*(1), 64–75. https://doi.org/10.1046/j.1523-1739.2000.99089.x

Oyer, E. B., Gries, G. A., & Rogers, B. J. (1959). The seasonal development of johnson grass plants. *Weeds*, *7*(1), 13. https://doi.org/10.2307/4040251

Padilla, R., Netto, S. L., & Da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.

Parico, A. I. B., & Ahamed, T. (2020). An aerial weed detection system for green onion crops using the You Only Look Once (YOLOv3) deep learning algorithm. *Engineering in Agriculture, Environment and Food*, *13*(2), 42-48.

Rauschert, E. S. J., Mortensen, D. A., Bjørnstad, O. N., Nord, A. N., & Peskin, N. (2009). Slow spread of the aggressive invader, *Microstegium vimineum* (Japanese stiltgrass). *Biological Invasions*, *12*(3), 563–579.

https://doi.org/10.1007/s10530-009-9463-y

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://dx.doi.org/10.1109/cvpr.2016.91

Ree, R. van der, Smith, D. J., & Grilo, C. (2015). *Handbook of Road Ecology*. John Wiley & Sons.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://dx.doi.org/10.1109/cvpr.2017.690

Reichard, S. H., & Hamilton, C. W. (1997). Predicting invasions of woody plants introduced into North America: Predicción de Invasiones de Plantas Leñosas Introducidas a Norteamérica. *Conservation Biology*, *11*(1), 193-203.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

Richardson, D. M., Pyšek, P., Rejmanek, M., Barbour, M. G., Panetta, F. D., & West, C. J. (2000). Naturalization and invasion of alien plants: concepts and definitions. *Diversity and distributions*, *6*(2), 93-107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

Riitters, K. H., & Wickham, J. D. (2003). How far to the nearest road? *Frontiers in Ecology and the Environment*, *1*(3), 125–129.

Ringland, J., Bohm, M., & Baek, S. R. (2019). Characterization of food cultivation along

roadside transects with Google Street View imagery and deep learning. *Computers and electronics in agriculture*, *158*, 36-50.

Rousselet, J., Imbert, C.-E., Dekri, A., Garcia, J., Goussard, F., Vincent, B., Denux, O., Robinet, C., Dorkeld, F., Roques, A., & Rossi, J.-P. (2013). Assessing species distribution using google street view: A pilot study with the pine processionary moth. *PLoS ONE*, *8*(10), e74918. https://doi.org/10.1371/journal.pone.0074918

Shuster, W. D., Herms, C. P., Frey, M. N., Doohan, D. J., & Cardina, J. (2005). Comparison of survey methods for an invasive plant at the subwatershed level. *Biological Invasions*, *7*(3), 393–403. https://doi.org/10.1007/s10530-004-3904-4

Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, *52*(3), 397-399.

Taaffe, E. J., Gauthier, H. L., & O'Kelly, M. E. (1996). *Geography of transportation*. MORTON O'KELLY.

The MathWorks, I. (2019). *Computer Vision Toolbox*. Natick, Massachusetts, United State. Retrieved from https://www.mathworks.com/help/vision/

The MathWorks, I. (2019). *Deep Learning Toolbox*. Natick, Massachusetts, United State. Retrieved from https://www.mathworks.com/help/deeplearning/

The MathWorks, I. (2019). *Image Processing Toolbox*. Natick, Massachusetts, United State. Retrieved from https://www.mathworks.com/help/images/

United States Forest Service. (n.d.). *Invasive species*. Retrieved June 4, 2022, from https://www.fs.fed.us/foresthealth/protecting-forest/invasive-species/

University of California Statewide IPM Program. (n.d.). *What is integrated pest management (IPM)? / UC Statewide IPM Program (UC IPM)*. Retrieved March

10, 2022, from https://www2.ipm.ucanr.edu/what-is-IPM/?src=redirect2refresh

US Census Bureau. (2021). *TIGER/Line geodatabases*. Census.Gov. Retrieved October 20, 2021, from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.2021.html#list-tab-AGZ2ZC2D5ZN46GFI00

Visser, V., Langdon, B., Pauchard, A., & Richardson, D. M. (2013). Unlocking the potential of Google Earth as a tool in invasion science. *Biological Invasions*, *16*(3), 513–534. https://doi.org/10.1007/s10530-013-0604-y

von der Lippe, M., & Kowarik, I. (2007). Long-Distance dispersal of plants by vehicles as a driver of plant invasions. *Conservation Biology*, *21*(4), 986–996. https://doi.org/10.1111/j.1523-1739.2007.00722.x

Wang, A., Zhang, W., & Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, *158*, 226–240. https://doi.org/10.1016/j.compag.2019.02.005

Wang, Q., Cheng, M., Huang, S., Cai, Z., Zhang, J., & Yuan, H. (2022). A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* Dunal seedlings. *Computers and Electronics in Agriculture*, *199*, 107194. https://doi.org/10.1016/j.compag.2022.107194

Werle, R., Jhala, A. J., Yerka, M. K., Anita Dille, J., & Lindquist, J. L. (2016). Distribution of herbicide-resistant shattercane and johnsongrass populations in sorghum production areas of Nebraska and northern Kansas. *Agronomy Journal*, *108*(1),

321-328.

Yan, Y., & Ryu, Y. (2021). Exploring Google Street View with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *171*, 278–296. https://doi.org/10.1016/j.isprsjprs.2020.11.022

Zheng, Y.-Y., Kong, J.-L., Jin, X.-B., Wang, X.-Y., & Zuo, M. (2019). CropDeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors*, *19*(5), 1058. https://doi.org/10.3390/s19051058

Zwaenepoel, A., Roovers, P., & Hermy, M. (2006). Motor vehicles as vectors of plant species from road verges in a suburban environment. *Basic and Applied Ecology*, *7*(1), 83–93. https://doi.org/10.1016/j.baae.2005.04.003