The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)

Accelerating Ecological Sciences from Above: Spatial Contrastive Learning for Remote Sensing

Johan Bjorck [1*], Brendan H. Rappazzo [1*], Qinru Shi [1], Carrie Brown-Lima [2]
Jennifer Dean [3], Angela Fuller [2 4 5], Carla Gomes [1]
[1] Department of Computer Science, Cornell University
[2] Department of Natural Resources and the Environment, Cornell University
[3] New York Natural Heritage Program, College of Environmental Science and Forestry, SUNY
[4] U.S. Geological Survey
[5] New York Cooperative Fish and Wildlife Research Unit
{njb225, bhr54, qs63, cjb37}@cornell.edu, jennifer.dean@dec.ny.gov, angela.fuller@cornell.edu, gomes@cs.cornell.edu

Abstract

The rise of neural networks has opened the door for automatic analysis of remote sensing data. A challenge to using this machinery for computational sustainability is the necessity of massive labeled data sets, which can be cost-prohibitive for many non-profit organizations. The primary motivation for this work is one such problem; the efficient management of invasive species – invading flora and fauna that are estimated to cause damages in the billions of dollars annually. As an ongoing collaboration with the New York Natural Heritage Program, we consider the use of unsupervised deep learning techniques for dimensionality reduction of remote sensing images, which can reduce sample complexity for downstream tasks and decreases the need for large labeled data sets. We consider spatially augmenting contrastive learning by training neural networks to correctly classify two nearby patches of a landscape as such. We demonstrate that this approach improves upon previous methods and naive classification for a large-scale data set of remote sensing images derived from invasive species observations obtained over 30 years. Additionally, we simulate deployment in the field via active learning and evaluate this method on another important challenge in computational sustainability – landcover classification – and again find that it outperforms previous baselines.

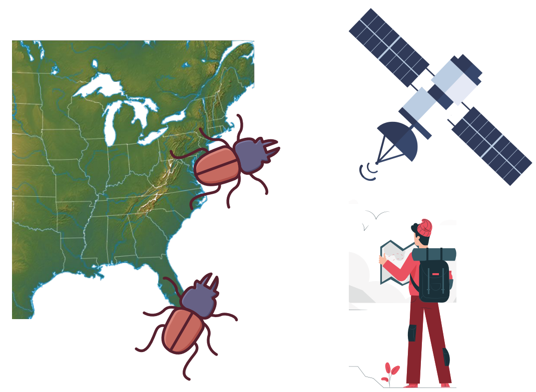

Figure 1: Invasive species management requires sending out observers to suitable locations across a large landscape. To effectively use limited resources it is imperative to send observers to the most important locations. In this work, we consider using unsupervised machine learning on remote sensing data to aid these efforts, using deep embedding techniques to improve sample complexity of species classification.

Introduction

In recent years, neural networks have made impressive strides in their potency and can now accurately predict faces (Ye et al. 2020) and translate natural language (Devlin et al. 2018). Besides progress in network architecture and optimization, this progress has been driven by large data sets such as Imagenet (Deng et al. 2009) and hugely parallel accelerators such as graphics processing units (GPUs) or tensor processing units (TPUs) (Jouppi et al. 2017). Besides these advances in machine learning, another field that has improved with more efficient data processing is remote sensing. While remote sensing via satellites has been used since the cold war, the first commercial satellite (IKONOS) was only launched in 1999. Researchers have been increasingly interested in applications of remote sensing to environmental questions (Jensen 2009; Lentile et al. 2006).

Machine learning for problems in computational sustainability is an active area of research (Gholami et al. 2019; Chen et al. 2016; Xue et al. 2017; Xie et al. 2015), and one problem that stands to benefit from both access to high-quality remote sensing data and machine learning methods is invasive species management. As part of an ongoing collaboration with the New York Natural Heritage Program, which manages and coordinates invasive species data in the state of New York (Department of Environmental Conservation 2018), this work considers remote sensing data for invasive species management. As part of the ImapInvasives

*Equal contribution

project (NatureServe 2020), the New York Natural Heritage Program collects data of observations across the state, and their current database includes over 200,000 observations of invasive species spread over 30 years. Collecting these data is laborious and requires sending out professionals or volunteer citizen scientists to conduct field surveys. To effectively utilize available workers, it is paramount to send them to suitable locations that have a high likelihood of containing invasive species. While exhaustively searching the state is impossible, the task is made easier by the fact that suitable habitats of various invasive species are often known, e.g., the hemlock wooly adelgid lives in coniferous hemlock forests (Holmes, Murphy, and Royle 2005). In practice, the allocation of observers to actual locations is often made by ecological experts.

Towards automating the task of deciding suitable locations for observations, we consider the task of predicting invasive species' locations from satellite images, see Figure 1 for a schematic and further description. A central problem with this approach is sample complexity; neural networks often require large labeled data sets, whereas many species might have few observations. Specifically, in this setting, closely monitoring such species before large outbreaks (meaning few observations) is ecologically important. However, in this setting, satellite imagery is easy to obtain, which suggests the use of unsupervised learning. With this in mind, we consider augmenting contrastive learning (Oord, Li, and Vinyals 2018) by utilizing the spatial structure of remote sensing data; training a neural network to classify nearby but non-identical satellite images as such. This naturally induces the network to generate low-dimensional embeddings, which can later be used for tasks like classification or active learning. As we demonstrate, this improves sample complexity over supervised methods and also is an improvement over previous methods of unsupervised learning of remote sensing images. In addition to evaluating our method on satellite images geo-referenced to an invasive species data set from New York Natural Heritage Program, we also consider using our method for another important problem in computational sustainability – landcover classification. We here consider the publicly available data sets Eurosat (Helber et al. 2019) and the U.S. Department of Agriculture's National Agricultural Imagery Program (NAIP) (Jean et al. 2019), and again show that our method beats baselines.

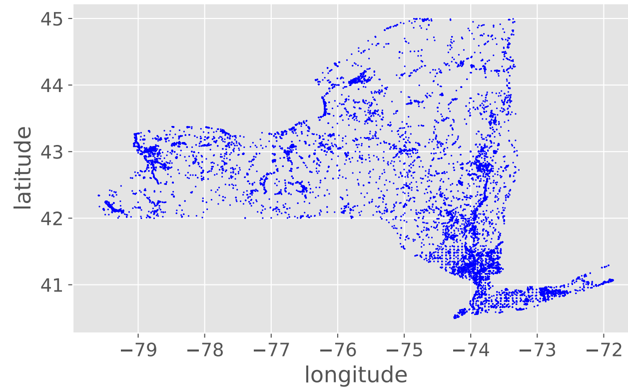

Figure 2: Our data set of invasive species observations covers the state of New York, spanning over 200 species and 30 years. Each dot corresponds to a unique observation.

Lastly, we simulate field deployment of our method via active learning and propose to perform active learning in the latent space of images, showing that it can improve upon traditional active learning. We summarize our contributions as follows.

- We introduce a new data set of remote sensing images for invasive species management, where images correspond to observations on the ground.
- We consider spatially augmented contrastive learning for remote sensing data as a method to improve sample complexity, and consistently find that it outperforms baselines across three data sets.
- We simulate field deployment of the method, proposing active learning in the *latent* space of satellite images.

Remote Sensing and Invasive Species

In this work we consider invasive species that are non-native and that cause some type of harm to the environment, economy, or human health. Examples include zebra mussels invading United States (US) freshwater bodies (Nienhuis, Haxton, and Dunkley 2014). It has been estimated that invasive species cause damages in the billions of dollars annually, just in the US (Pimentel, Zuniga, and Morri-

Species	Description	Observations	% of dataset
Water Chestnut	Floating aquatic plant that hinders boats and crowds out native plants	285	4.39
Honeysuckle	Terrestrial plants that form monotypic stands and reduces diversity	295	4.54
Oriental Bittersweet	Woody vine that smothers and uproot trees	307	4.72
Japanese Knotweed	Terrestrial plant that forms monotypic stands and reduces diversity	1287	19.81
Garlic Mustard	Terrestrial plant that forms monotypic stands and reduces diversity	653	10.05
Japanese Barberry	Terrestrial plant that forms monotypic stands and reduces diversity	459	7.06
Common Reed Grass	Terrestrial plant that forms monotypic stands and reduces diversity	1174	18.07
Purple Loosestrife	Shoreline plant that clogs waterways and reduces wetland habitat	918	14.13
EurasianWater-milfoil	Submerged aquatic plant that hinders boats and crowds out native plants	441	6.79
Multiflora Rose	Terrestrial plant that forms monotypic stands and reduces diversity	679	10.45

Table 1: Class names and distribution for the invasive species data set, as well as a description of their ecological relevance.

son 2005). A famous example is the hemlock woolly adelgid, which initially came to the US from Japan (Oten et al. 2014). The insect feeds on the sap at the base of hemlock needles, disrupting nutrient flow and eventually killing the tree. Due to the ecological importance of hemlocks in many forest ecosystems, researchers across the US are working on finding efficient strategies to monitor, mitigate, and eradicate the hemlock woolly adelgid as well as all other invasive species. As part of an ongoing effort in the state of New York to monitor invasive species, the New York Natural Heritage Program uses iMapInvasives (NatureServe 2020) to collect and synthesize invasive species data across the state going back more than 30 years (NatureServe 2020). The state is divided into eight invasive species regions, each with partnerships which monitor their region using a combination of paid employees and citizen scientists. Records of observed invasive species are reported to iMapInvasives as the central database. The database currently consists of over 200,000 individual observations, each containing a location and time, the species found, the observer's name, etc. Figure 2 illustrates the geographical spread of recorded observations. We will consider the ten most observed invasive species, listed in table 1, and will construct a remote sensing data set from these observations to be used for downstream tasks. On the fine spatial scale, observations are strongly correlated, as it is typical for one observer to observe multiple invasive species some meters away from another observer. Additionally, some locations have more observations than others, such as data near large cities. To make the data set approximately spatially balanced, we randomly sub-sample the observations across a grid. We divide the state of New York into a grid corresponding to 0.01 degrees latitude and longitude, and only select one observation per square in this grid, and further make sure that there are no neighboring (horizontally, vertically or diagonally) observations. This results in a data set of 6498 observations, and ten classes – corresponding to unique species – that are roughly balanced between species. For this work we do not consider any temporal information about the observations, such as what date or time an observation was made. See Table 1 for further details. We then obtain 512x512 pixels red, green, blue (RGB) remote sensing images corresponding to these locations via Google Maps, see Figure 4 for examples. While most invasive species cannot be seen from satellite, their tendency to prefer certain cover types, e.g., the Hemlock Woolly Adelgid prefers hemlock trees found in coniferous forest, will be useful as ecosystem traits can be observed via satellites. Given this data set, we first consider using unsupervised machine learning to generate low-dimensional embeddings that efficiently allows us to classify what invasive species inhabit what regions based upon historical data. The ultimate goal of this line of work is to use machine learning predictions to actually decide what pieces of land are susceptible to invasive species. Later in the paper, we simulate this by considering an active learning approach on this historical data set and defer field deployment to future work.

Embeddings

Given a remote sensing image x we wish to be able to generate low-dimensional embeddings $y = f(x)$ for some mapping f. The perhaps most well-known applications of embeddings are so-called word-embeddings, where individual words are represented by dense vectors suitable for neural network computation (Mikolov et al. 2013). Given some corpora of text D, one initializes each word w in a language to be represented by some vector $v(w)$ and then obtains the final word embeddings as the solution to some optimization problem. Typically, one considers some loss function ℓ using the word w and its neighbor n, i.e., we have

$$\min \mathbb{E}_{w,n\sim D}\Big[\ell\big(v(w), v(n)\big)\Big]$$

It is often desirable to choose the loss function ℓ such that words that are used in similar contexts, i.e., have similar neighbors, have similar embeddings. The motivation for defining the loss function in this manner can be motivated by the J.R. Firth quote, "You shall know a word by the company it keeps". We consider an embedding for a satellite image x, but whereas there is a discrete fixed number of

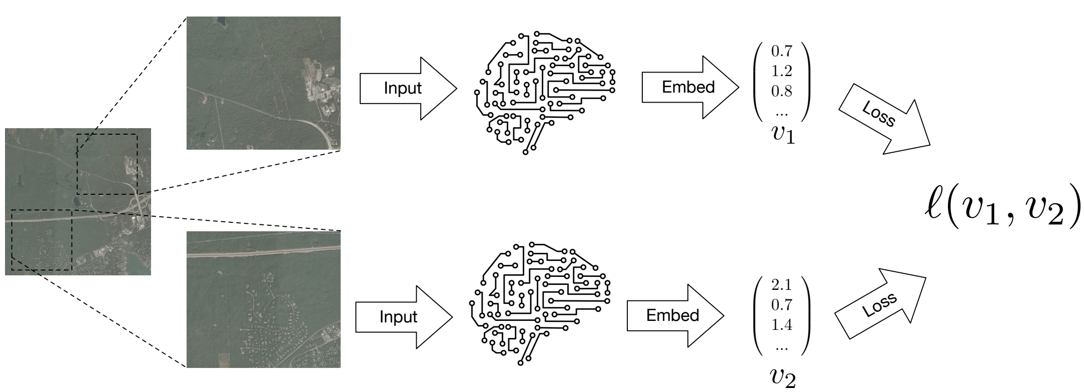

Figure 3: Given a remote sensing landscape, we consider two patches close to each other. These images are then fed into the same neural network which generate embeddings v_1, v_2 of the images. Given a collection of such embeddings, we want to be able to classify neighbours as such, and use the inner product $v_1^T v_2$ as the logit.

words in a language, we will let the embeddings be given by a neural network f. Inspired by this strategy of considering close words, it is natural to apply the same idea that the embeddings of satellite images of nearby locations should have similar embeddings, see Tile2Vec or Patch2Vec (Jean et al. 2019; Fried, Avidan, and Cohen-Or 2017) that rely on triplet loss. Instead of directly optimizing embeddings via the triplet loss, we obtain them as a byproduct of a classification task, extending contrastive learning (Oord, Li, and Vinyals 2018; Bachman, Hjelm, and Buchwalter 2019) to spatial domains by utilizing the neighborhood relationship induced by the spatial distribution of remote sensing images. Specifically, given two remote sensing images x, n, where n is a neighbor to x, we train the neural network f to generate embeddings that allow us to conclude that x and n indeed are neighbors. For an illustration, see Figure 3. A simple strategy is to cast this as a classification problem and use the softmax loss. If the embeddings are column vectors, we treat their inner product as the actual logit and then consider the soft-max cross-entropy loss.

$$\ell(x, n) = -\log \frac{\exp\big(f(x)^T f(n)\big)}{\sum_j \exp\big(f(x)^T f(j)\big)} \tag{1}$$

In practice, computing the denominator is expensive, and one can approximate it by only considering negative examples from the same batch. For a schematic illustration of the method, see Figure 3. One can further enlarge the data set by considering augmentations such as, rotations that the natural landscape is approximately invariant under. It has been observed that one can slightly improve contrastive learning by scaling the logits by some fixed parameter T and not using the embeddings of the final layer, but instead adding a small head multi-layer perceptron (MLP) on top of the convolutional neural network (CNN) for training but then using intermediate representations from the CNNs as representations (Chen et al. 2020).

	Invasive Data Set	
Embeddings	RFC	LR
Contrastive	25.63 ± 1.54	26.71 ± 1.71
Tile2Vec	23.07 ± 1.03	24.16 ± 1.57
AutoEncoder	22.50 ± 0.78	21.91 ± 1.14
PCA	22.16 ± 0.80	22.36 ± 1.37
ICA	22.24 ± 1.11	19.55 ± 1.23

Table 2: Accuracy for unsupervised setting. All experiments were run for 10 rounds, and the average value is given $\pm$ the standard deviation.

Experiments

In this section, we primarily focus on our invasive species data set to evaluate the feature extraction from unlabeled remote sensing images. The dataset is constructed as per previous section. We also perform active learning experiments to simulate deploying our method in the field and additionally perform experiments for two external remote sensing data sets.

Unsupervised Experiments

We first evaluate whether the embeddings generated via our methods are useful for classification, comparing to Tile2vec (Jean et al. 2019) and some further baselines which we describe here. **Tile2Vec** uses a triplet loss (Hoffer and Ailon 2015) to train a feature extracting network to push geographically nearby tiles close together in the extracted feature space. The Tile2Vec and contrastive feature extractors are built with the ResNet-18 architecture (He et al. 2015), with the last layer set to have 256 neurons/features. The contrastive method also uses a ResNet-18 architecture, plus

Figure 4: Examples images from all three data sets. The invasive species data set correspond to observations of invasive species (given in table 1) across the state of New York. The Eurosat data set corresponds to satellite images over ten types of landcovers across continental Europe (Helber et al. 2019). The NAIP data set corresponds to images from California obtained via the national agriculture imaging program (Jean et al. 2019).

a two-layer top module with 256-neurons for embedding which is discarded after training (i.e. features are obtained from the underlying ResNet) (Chen et al. 2020). The models are trained for 150 epochs with a batch size of 256 and a learning rate of 0.1, both using the Adam optimizer (Kingma and Ba 2014). Tile2Vec uses the triplet loss with a margin of 0.1 following (Jean et al. 2019), whereas the contrastive method is trained with the loss in eq. (1). See the Appendix for further hyperparameters. The **autoencoder** (Kramer 1991) baseline has an encoding module consisting of three convolutional layers with 8, 16 and 32 filters, respectively. This is followed by two fully connected layers of size 256 and 128, meaning the feature space has 128 dimensions. The decoding module had a single, fully connected layer of size 512, followed by three transpose convolutional layers. All convolutional layers were followed by a max-pooling layer, and all layers, except the output layer, were passed through a Leaky ReLu activation with a negative slope of 0.01. The autoencoder was optimized to minimize the mean squared error between the input image and reconstruction, training over 40 epochs with a batch size of 256 and using the Adam optimizer with a learning rate of 0.001. For principal component analysis (**PCA**) (Tipping and Bishop 1999) and independent component analysis (**ICA**) (Hyvärinen and Oja 2000) each image was flattened to be a vector of size 12,288. The top 10 principle or independent components are then computed, and the activations of these components for each image is treated as extracted features.

For all methods, the data were prepared in the same manner. The images are normalized to have zero mean and unit variance. We consider random 64x64 parts of the original 512x512 image, for Tile2Vec and contrastive two neighbors consists of two such parts from the same base image. In addition to using nearby parts of a remote sensing image, we also rotate and flip the images randomly and randomly zero out 16-by-16 sub-images to extend the data set further. The data set is first divided into a 70 percent set for unsupervised training of the feature extractors; the remaining 30 percent of the data is then randomly split into two 15 percent sets for training and evaluating the top-level classifier. We assess feature quality for a total of 10 rounds by evaluating the accuracy of a top layer classifier using the extracted features and ground truth labels – using either a logistic regression classifier (LR) or a random forest classifier (RFC) (Pedregosa et al. 2011). We chose these methods as they are well-known and often perform well in practice. The RFC is trained using 100 decision tree learners and Gini impurity as the criterion for splitting; see the Appendix for further hyperparameters. We report test accuracy for the top layer classifier using the extracted features, giving the mean and standard deviation accuracy for these methods in Table 2. As can be seen, our contrastive method outperforms all baselines for both classifiers. The predictions are likely to further improve with a larger data set, and we emphasize that the task is difficult as one cannot directly see invasive species from the images but must instead consider what habitat might be suitable for them.

Supervised Experiments

In ecological or biological domains, it is often easy to obtain unlabeled remote sensing data but generating labels requires sending domain experts to the field, which is a laborious process. A strong unsupervised feature extractor can potentially lead to a high accuracy classifier with much fewer labels, corresponding to substantial savings in fieldwork. With this in mind, we investigate our method's accuracy compared to fully supervised methods with different amounts of labeled data available. We consider strong baseline deep learning classifiers DenseNet (Huang, Liu, and Weinberger 2016), ResNet (He et al. 2015) and AlexNet (Krizhevsky, Sutskever, and Hinton 2012), and additionally compare to the features extracted via Tile2vec. The data set is split into a training set of size 70 percent and a testing data set composed of the remaining 30 percent. From the training set, a variable percentage of the labels were then removed. This was done to simulate a real-world setting where there are ample unlabeled data for unsupervised methods to use, but

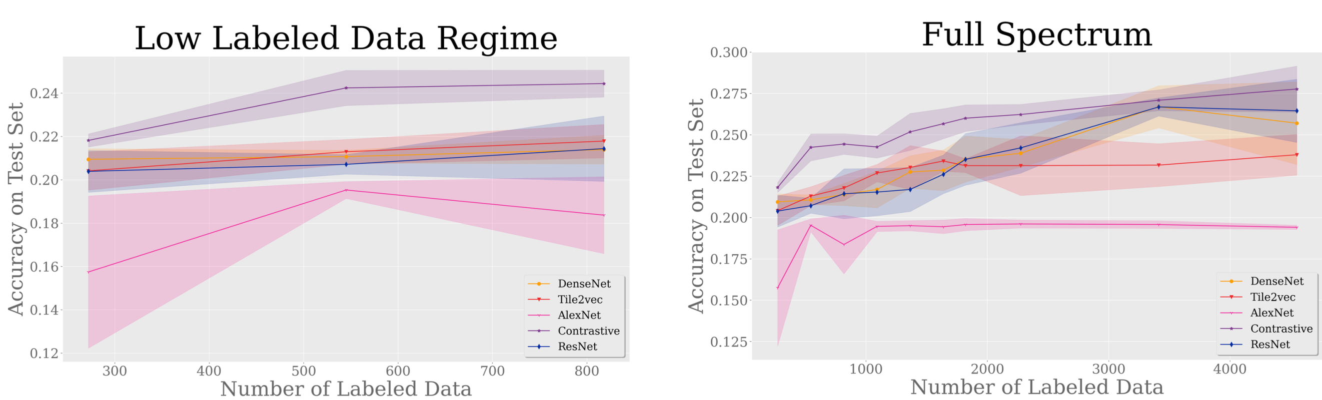

Figure 5: Available labeled data vs accuracy for supervised methods and unsupervised methods trained on all images (but not all labels). In this low-data regime (left) and zoomed out full spectrum of available labels (right), spatial contrastive learning outperforms classical supervised methods.

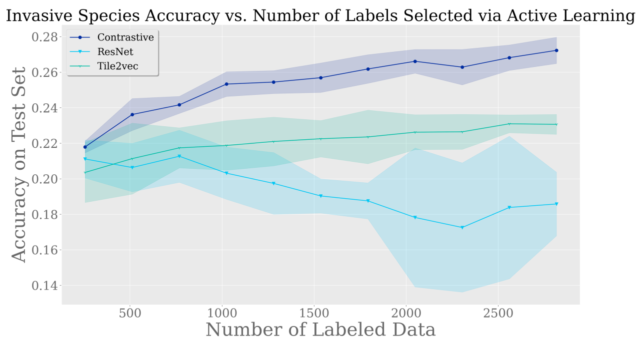

Figure 6: We simulate field deployment by performing active learning across the invasive species data set. Unlike classical active learning, where one queries for both images and labels, we propose to perform the active learning in the embedding space obtained from unsupervised models. This approach outperforms traditional active learning, and spatial contrastive learning outperforms Tile2vec.

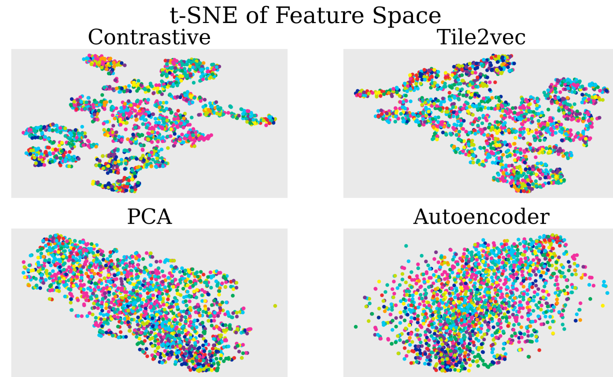

Figure 7: t-SNE visualization of the feature space for invasive species data set for our method, Tile2Vec, PCA and Autoencoder, where each color indicates a particular invasive species class. The illustration suggests that Tile2Vec encourage looser clusters that can lead to generalization error, which could be a reason our method performs better.

few labels for supervised methods. Within the training set, we ran experiments with the following percent of labeled data available: 6, 12, 18, 24, 30, 36, 40, 50, and 75 percent. For the unsupervised methods (contrastive and Tile2Vec), we first train the unsupervised feature extraction on the entire training data set, using the same hyperparameters as the unsupervised experiments. We then train a top-level classifier (RFC) on the available labeled data using the extracted features as input. For the fully supervised methods, we train them on the available labeled data for 40 epochs using the Adam optimizer with a learning rate of 0.1 to optimize the cross-entropy loss, resulting in convergence for the loss. We then test each model's accuracy on the test set. We repeat each of these experiments for five rounds and report the mean and standard deviation accuracy. As can be seen in Figure 5, our method outperforms all others. This highlights how our method can be used to greatly reduce the number of needed labels, and therefore the cost, to obtain an accurate classifier. With a larger amount of labeled data, fully supervised methods, like DenseNet or ResNet, likely would match the performance of our method, but the experiments suggest that when labeled data are scarce, spatial contrastive learning provides efficient feature extraction.

Active Learning

The ultimate goal of our collaboration with the New York Natural Heritage Program is to use remote sensing images to direct ecologists to locations deemed likely to contain invasive species. To roughly simulate this setting, we consider performing active learning over our invasive species data set. We are given a fixed number of queries for labels and must use these to train as accurate a prediction model as possible (evaluated on a held-out test-set). Unlike traditional active learning where one chooses both images and labels to add to the train set, we propose to use all images for unsupervised pre-training, and then only conduct active learning on the labels. In our setting, remote sensing data are inexpensive but sending ecologists to perform observations on the ground is more expensive. The idea of conducting active learning only on the labels has the potential to greatly speed up ecological work but is only possible if useful features can be extracted in an unsupervised fashion.

In this experimental setting, we compare our contrastive model, Tile2Vec and standard active learning (i.e. selecting both images and labels) using ResNet-18. The data are prepared in the same manner as the supervised experimental setting. We first train the contrastive method and Tile2Vec unsupervised feature extraction using the same setting as described in the unsupervised experiments. We then randomly initialize each method to have 256 labels and allow it to train a supervised model. For the contrastive method and Tile2Vec we trained a RFC from the features extracted to the labels, using 100 decision tree learners and the Gini purity as the splitting criterion. For ResNet, for each round the ResNet model was trained on all available labeled data for 40 epochs, with a batch size of 256, optimizing the cross-entropy loss with the Adam optimizer and a learning rate of 0.1. Then we run a series of 10 rounds of active learning, using the entropy sampling method for active learning (Settles 2012). Each round we generate predictions on all unlabeled images in the training set and take the set (of size 256) which produced the largest entropy in the classification predictions. This selected set is then added to the available labeled data for each method, and the model re-trains on the now larger train set and reports its accuracy on the test set. This experiment was run five times for the contrastive method, Tile2Vec, and ResNet. In Figure 6, we plot the mean and standard deviation accuracy against the amount of labeled data available. This demonstrates that our method can be used in an active learning setting to guide which labels should be taken; we hope to study this approach further in the future. This model is, of course, a simplification compared to actual field deployment, and many practical differences compared with real deployment remain.

EuroStat		
Embeddings	RFC	LR
Contrastive	71.47 ± 0.40	71.23 ± 0.67
Tile2Vec	60.49 ± 0.63	49.77 ± 0.56
AutoEncoder	60.22 ± 0.92	57.27 ± 0.51
PCA	65.72 ± 0.80	43.13 ± 0.94
ICA	65.30 ± 0.86	21.25 ± 5.11

Table 3: Accuracy for Eurosat (Helber et al. 2019). All experiments were for 10 rounds, and the average value is given ± the standard deviation.

NAIP		
Embeddings	RFC	LR
Contrastive	66.47 ± 1.88	58.38 ± 1.96
Tile2Vec	62.70 ± 1.51	53.17 ± 1.52
AutoEncoder	61.50 ± 2.77	40.65 ± 2.03
PCA	60.92 ± 1.43	57.55 ± 2.51
ICA	62.58 ± 1.97	34.08 ± 1.58

Table 4: Accuracy for NAIP (Jean et al. 2019). All experiments were for 10 rounds, and the average value is given ± the standard deviation.

Qualitative Analysis

To probe the learned features, we use t-SNE (Maaten and Hinton 2008) to visualize the features extracted by our method, Tile2vec, the autoencoder and PCA, as seen in Figure 7. For this experiment, the embedding data were extracted from each method for the test set after training each method to convergence, using the same parameters as per the unsupervised experimental setting. The illustration suggests that Tile2Vec and spatial contrastive learning results in a clearer structure than PCA and autoencoders. Further, we suspect that the L2 loss used in Tile2vec may not constrain the clusters as can be seen in this visualization, perhaps leading to a weakened ability for the model to generalize.

Additional Data Sets

Eurostat While the invasive species application is the main focus of this work, we conduct experiments on additional data sets to show the generality of the proposed method. We consider landcover classification, where one attempts to classify a remote sensing image as belonging to some specific landcover type (forest, road, river, etc.). This task has practical implications in computational sustainability and can, e.g., be used for monitoring deforestation. We first consider the Eurostat data set, which consists of 27,000 Sentinel-2 satellite images of various landcover types from Europe (Helber et al. 2019). The data and baselines were all prepared as for the invasive species data set, and the results of our experiments on this data set can be seen in Table 3. As can be seen, our method outperforms all other feature extractors on this data set.

NAIP We additionally consider the NAIP data set (Jean et al. 2019), which contains a fourth spectrum band, which highlights our method's ability to handle multi-spectral remote sensing. A difference in this experimental setting is that the NAIP data set has train and test set sources from different geographical locations and that one must obtain feature extraction that is robust under such distributional shift. See (Jean et al. 2019) for details. Again, for this data set, we consider the same unsupervised experimental setting as per the invasive species data set. We use the entire training set to train our unsupervised methods, and then split the test set into two equal sets. For the PCA and ICA feature extractors, because of the fourth color channel, the input vectors were of size 16,384 as opposed to 12,288; otherwise, the data and baselines were all prepared in the same manner. The results of our experiments on this data set can be seen in Table 4. As can be seen, our method outperforms all other feature extractors on this data set.

Challenges and Opportunities

Whereas this paper has focused on computational aspects of invasive species management, deploying and using our models has many practical challenges and opportunities that we here expand upon. Firstly it is important to note that even an accuracy around 25 % can be useful for directing fieldwork and that it can complement classical approaches. We also note that the problem is hard as we only observe the habitat and not the invasive species themselves. Certain habitats can be favorable for invasive species, but that does not necessarily imply species presence. We also emphasize that there are variations within invasive species habitats. Ecologists know that the hemlock wooly adelgid lives off of hemlock trees, but an exact understanding of how the forest characteristics interact with the spreading rate is lacking (Oten et al. 2014). Not all hemlock forests are identical, and there might be variations in e.g. tree density that influence spreading. Furthermore, land cover types are often coarse and might be on the level of "evergreen forest" rather than specifying e.g. tree species composition. Proximity to roads, trails, and water bodies can often impact invasive species spread, and their presence is easily detected from satellite images but not necessarily captured by land cover. The habitats can also pose a problem for our machine learning models. Many of the terrestrial plants we used can occur in the same or very similar forested habitats (and same for the aquatic plants in aquatic habitats). This could result in misclassifications in machine learning outputs for a particular species. Misclassification could have cost implications for managers by either sending managers to unsuitable sites or possibly missing a key population that should be managed.

Secondly, we highlight how the data are collected. The New York State (NYS) invasive species program is managed by a collection of regional organizations which use paid professionals and citizen scientists. Strategies include recruiting citizen scientists for shorter fieldwork excursions, allowing citizens to report invasive species via an online reporting tool, or inviting the public to participate in plant removal events. How the data are collected likely leads to some bias, for example, locations that are easier to reach might be monitored more frequently. Bias is common in citizen science applications, and e.g. the eBird project suffers from roadside bias (Chen and Gomes 2019). However, we note that many invasive species spread via humans, so bias towards populated areas is not necessarily bad.

Related Work

Unsupervised deep learning has a long history (Kramer 1991). A popular line of work employs auto-encoders (Hinton and Zemel 1994). Contrastive predictive coding has also been researched since (Oord, Li, and Vinyals 2018), and typically relies on predicting parts of the input given other parts (Bachman, Hjelm, and Buchwalter 2019; Srinivas, Laskin, and Abbeel 2020). Specifically, in (Oord, Li, and Vinyals 2018) the method relies on finely dividing natural images into subparts and then autoregressively making predictions. The idea of contrastive coding has inspired a lot of recent work (Chen et al. 2020; He et al. 2020). We spatially augment contrastive learning methods, and instead of considering crops of the same image, we use non-overlapping parts of the same landscape – relying on spatial smoothness of landscape features. The strategy of considering neighbors is popular in natural language processing (NLP) (Devlin et al. 2018), and is also used for word embeddings (Mikolov et al. 2013; Pennington, Socher, and Manning 2014). With the advent of deep learning, machine learning for remote sensing has received much attention. The most closely related work is Tile2Vec (Jean et al. 2019), which uses a strategy reminiscent of word embeddings to generate remote sensing embeddings, specifically using the triplet loss. The work of (Fried, Avidan, and Cohen-Or 2017) also uses the triplet loss for geographic data but instead relies on supervision. Contemporaneous work also includes (Kang et al. 2020), which similarly to this study considers unsupervised learning for remote sensing, we emphasize that our work also includes applications to active learning. An important application of remote sensing is poverty mapping, where given access to remote sensing data, one tries to predict economic conditions "on the ground" (Xie et al. 2015). Another important use case is the prediction of crop yield from remote sensing data (Setiyono, Nelson, and Holecz 2014; Wang et al. 2018). Invasive species management is an important ecological problem with economic implications, and computational aspects of the problem have received considerable interest. Researchers have used remote sensing via airplanes to identify invasive species from handcrafted features (Ustin et al. 2002; Asner et al. 2008; Piiroinen et al. 2018). Researchers have used reinforcement learning (Taleghan et al. 2015), mixed integer programming (Büyüktahtakın, Feng, and Szidarovszky 2014) and stochastic dynamic programming (Shea and Possingham 2000) to generate management strategies. Modelling work includes Hawkes processes (Gupta et al. 2018), extensions of the firefighter problem (Spencer 2012) and predator-prey dynamics (Bjorck et al. 2018).

Conclusions

In this work, we have considered the use of remote sensing data for invasive species management, motivated by an ongoing collaboration with the New York Natural Heritage Program. By spatially augmenting contrastive coding methods, we show how to obtain low-dimensional embeddings of remote sensing data. Our experiments show that this method outperforms baselines, and we additionally show how one can perform active learning in this embedding space to improve sample complexity. For future work, we hope to further study how to integrate these methods into deployment.

Acknowledgements

We thank Rich Bernstein for technical support. This research is supported in part by CNS-0832782 (Expeditions). This material is based upon work supported by the National Science Foundation under Grant Number CCF-1522054. This material is also based upon work supported by the Air Force Office of Scientific Research under award number FA9550-18-1-0136. We are also grateful from generous support from the TTS foundation. We are also thankful for funding from the New York State Environmental Protection Fund as administered by the New York State Department of Environmental Conservation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Ethics Statement

Our primary motivation for this work is using remote sensing data to accelerate invasive species management. There is a long history of using remote sensing data for military purposes, and we see the case for advancing the use of these data to improve unsupervised learning and monitoring of natural resources and the environment including not only invasive species but other examples such as monitoring deforestation, fishing, and pollution levels.

References

Asner, G. P.; Knapp, D. E.; Kennedy-Bowdoin, T.; Jones, M. O.; Martin, R. E.; Boardman, J.; and Hughes, R. F. 2008. Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and LiDAR. *Remote Sensing of Environment* 112(5): 1942–1955.

Bachman, P.; Hjelm, R. D.; and Buchwalter, W. 2019. Learning representations by maximizing mutual information across views. In *Advances in Neural Information Processing Systems*, 15535–15545.

Bjorck, J.; Bai, Y.; Wu, X.; Xue, Y.; Whitmore, M.; and Gomes, C. 2018. Scalable Relaxations of Sparse Packing Constraints: Optimal Biocontrol in Predator-Prey Networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Büyüktahtakın, İ. E.; Feng, Z.; and Szidarovszky, F. 2014. A multi-objective optimization approach for invasive species control. *Journal of the Operational Research Society* 65(11): 1625–1635.

Chen, D.; and Gomes, C. P. 2019. Bias reduction via end-to-end shift learning: Application to citizen science. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 493–500.

Chen, Q.; Song, X.; Yamada, H.; and Shibasaki, R. 2016. Learning deep representation from big and heterogeneous data for traffic accident inference. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709* .

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Department of Environmental Conservation, N. 2018. *New York State Invasive Species Comprehensive Management Plan*. URL https://www.dec.ny.gov/docs/lands_forests_pdf/iscmpfinal.pdf.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Fried, O.; Avidan, S.; and Cohen-Or, D. 2017. Patch2vec: Globally consistent image patch representation. In *Computer Graphics Forum*, volume 36, 183–194. Wiley Online Library.

Gholami, S.; Xu, L.; Mc Carthy, S.; Dilkina, B.; Plumptre, A.; Tambe, M.; Singh, R.; Nsubuga, M.; Mabonga, J.; Driciru, M.; et al. 2019. Stay Ahead of Poachers: Illegal Wildlife Poaching Prediction and Patrol Planning Under Uncertainty with Field Test Evaluations. *arXiv preprint arXiv:1903.06669* .

Gupta, A.; Farajtabar, M.; Dilkina, B.; and Zha, H. 2018. Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management. In *IJCAI*, 3385–3392.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385. URL http://arxiv.org/abs/1512.03385.

Helber, P.; Bischke, B.; Dengel, A.; and Borth, D. 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(7): 2217–2226.

Hinton, G. E.; and Zemel, R. S. 1994. Autoencoders, minimum description length and Helmholtz free energy. In *Advances in neural information processing systems*, 3–10.

Hoffer, E.; and Ailon, N. 2015. Deep Metric Learning Using Triplet Network. In Feragen, A.; Pelillo, M.; and Loog, M., eds., *Similarity-Based Pattern Recognition*, 84–92. Cham: Springer International Publishing. ISBN 978-3-319-24261-3.

Holmes, T. P.; Murphy, E. A.; and Royle, D. D. 2005. The economic impacts of hemlock woolly adelgid on residential landscape values: Sparta, New Jersey case study. In *In: Onken, B.; Reardon, R. comps., eds. Proceedings of the 3rd symposium on hemlock woolly adelgid in the eastern United States, FHTET-2005-01. Morgantown, WV: US Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team: 15-24.*

Huang, G.; Liu, Z.; and Weinberger, K. Q. 2016. Densely Connected Convolutional Networks. *CoRR* abs/1608.06993. URL http://arxiv.org/abs/1608.06993.

Hyvärinen, A.; and Oja, E. 2000. Independent component analysis: algorithms and applications. *Neural Networks* 13(4): 411 – 430. ISSN 0893-6080. doi:https://doi.org/10.1016/S0893-6080(00)00026-5. URL http://www.sciencedirect.com/science/article/pii/S0893608000000265.

Jean, N.; Wang, S.; Samar, A.; Azzari, G.; Lobell, D.; and Ermon, S. 2019. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3967–3974.

Jensen, J. R. 2009. *Remote sensing of the environment: An earth resource perspective 2/e*. Pearson Education India.

Jouppi, N. P.; Young, C.; Patil, N.; Patterson, D.; Agrawal, G.; Bajwa, R.; Bates, S.; Bhatia, S.; Boden, N.; Borchers, A.; et al. 2017. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 1–12.

Kang, J.; Fernandez-Beltran, R.; Duan, P.; Liu, S.; and Plaza, A. J. 2020. Deep unsupervised embedding for remotely sensed images based on spatially augmented momentum contrast. *IEEE Transactions on Geoscience and Remote Sensing* .

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Kramer, M. A. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal* 37(2): 233–243.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 25*, 1097–1105. Curran Associates, Inc. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Lentile, L. B.; Holden, Z. A.; Smith, A. M.; Falkowski, M. J.; Hudak, A. T.; Morgan, P.; Lewis, S. A.; Gessler, P. E.; and Benson, N. C. 2006. Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire* 15(3): 319–345.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9(Nov): 2579–2605.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

NatureServe. 2020. *iMapInvasives: NatureServe's online data system supporting strategic invasive species management*. URL http://www.imapinvasives.org.

Nienhuis, S.; Haxton, T. J.; and Dunkley, T. C. 2014. An empirical analysis of the consequences of zebra mussel invasions on fisheries in inland, freshwater lakes in Southern Ontario. *Management of Biological Invasions* 5(3): 287.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* .

Oten, K. L.; Merkle, S. A.; Jetton, R. M.; Smith, B. C.; Talley, M. E.; and Hain, F. P. 2014. Understanding and developing resistance in hemlocks to the hemlock woolly adelgid. *Southeastern Naturalist* 13(6): 147–167.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825–2830.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Piiroinen, R.; Fassnacht, F. E.; Heiskanen, J.; Maeda, E.; Mack, B.; and Pellikka, P. 2018. Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification. *Remote Sensing of Environment* 218: 119–131.

Pimentel, D.; Zuniga, R.; and Morrison, D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics* 52(3): 273–288.

Setiyono, T.; Nelson, A.; and Holecz, F. 2014. Remote sensing-based crop yield monitoring and forecasting. *Crop monitoring for improved food security* .

Settles, B. 2012. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning Series. Morgan & Claypool. ISBN 978-1-60845-725-0. URL https://books.google.com/books?id=z7toC3z_QjQC.

Shea, K.; and Possingham, H. P. 2000. Optimal release strategies for biological control agents: an application of stochastic dynamic programming to population management. *Journal of Applied ecology* 37(1): 77–86.

Spencer, G. 2012. Robust cuts over time: Combatting the spread of invasive species with unreliable biological control. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Srinivas, A.; Laskin, M.; and Abbeel, P. 2020. Curl: Contrastive unsupervised representations for reinforcement learning. *arXiv preprint arXiv:2004.04136* .

Taleghan, M. A.; Dietterich, T. G.; Crowley, M.; Hall, K.; and Albers, H. J. 2015. PAC optimal MDP planning with application to invasive species management. *The Journal of Machine Learning Research* 16(1): 3877–3903.

Tipping, M. E.; and Bishop, C. M. 1999. Probabilistic Principal Component Analysis 13.

Ustin, S.; DiPietro, D.; Olmstead, K.; Underwood, E.; and Scheer, G. 2002. Hyperspectral remote sensing for invasive species detection and mapping. volume 3, 1658 – 1660 vol.3. ISBN 0-7803-7536-X. doi:10.1109/IGARSS.2002.1026212.

Wang, A. X.; Tran, C.; Desai, N.; Lobell, D.; and Ermon, S. 2018. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–5.

Xie, M.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2015. Transfer learning from deep features for remote sensing and poverty mapping. *arXiv preprint arXiv:1510.00098* .

Xue, Y.; Wu, X.; Morin, D.; Dilkina, B.; Fuller, A.; Royle, J. A.; and Gomes, C. P. 2017. Dynamic Optimization of Landscape Connectivity Embedding Spatial-Capture-Recapture Information. In *AAAI*, 4552–4558.

Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; and Hoi, S. C. 2020. Deep learning for person re-identification: A survey and outlook. *arXiv preprint arXiv:2001.04193* .