Implicit Land Use Mapping Using Social Media Imagery

Connor Greenwell
University of Kentucky

Scott Workman
DZYNE Technologies

Nathan Jacobs
University of Kentucky

***Abstract*—Land use classification is a central remote sensing task with a broad range of applications. Typically this is represented as a supervised learning problem, the first step of which is to develop a taxonomy of discrete labels. However, such categories are restricted in the range of uses they can convey and arbitrary decisions are often required when defining the categories. Instead, we argue that the abstract notion of land use can be indirectly characterized by the types and quantities of common objects found in an area. To capture the presence of such objects, we propose an implicit approach to defining and estimating land use that relies on sparsely distributed social media imagery but retains the benefits of dense coverage provided by satellite imagery. Our method is formulated as a convolutional neural network that operates on satellite imagery and outputs a probability distribution over quantities of objects common in social media imagery at that location. We show that the learned feature representation is discriminative for existing land use categories.**

***Index Terms*—multi-task learning, weak supervision, semantic transfer, data fusion**

I. Introduction

Accurately estimating land use (and equivalently land cover), and how it changes over time, is critical for a wide variety of applications, including: disaster response, famine forecasting, population density mapping, risk hazard assessment, and monitoring refugee camps for human rights violations. Given the prohibitive cost of manual field assessments, many methods have explored how to characterize properties of land use directly from satellite imagery.

Typically, approaches for automatic land use classification [1], [2] formulate the problem as a supervised learning task and follow a standard set of steps. First, a taxonomy of discrete target labels is determined. Second, satellite imagery is annotated with these labels by human experts, either in the field or by inspecting the imagery. Finally, these annotations are used to train an image classification model via supervised learning. The resulting model can then be applied to label new imagery as it is collected.

This approach depends on developing a meaningful taxonomy of target labels, which often requires making arbitrary decisions. Often, a large factor when determining the final label taxonomy is the difficulty of obtaining annotated training data. This means the resulting set of labels are not necessarily ideal for a given application. This is especially true when using a small number of categories.

978-1-7281-4732-1/19/$31.00 2019 IEEE

Our goal is to implicitly characterize land use by modeling the types and quantities of objects that are likely to be found at a given location. In other words, we want to remove the manual first step of defining a label taxonomy, which can lead to loss of information. Ground-level imagery contains rich information about the setting in which it was captured. Additionally, modern deep-learning based computer vision techniques can be used to automatically extract this information. For this reason, we follow a weakly supervised approach that takes advantage of sparsely located geotagged ground-level imagery as the source of target labels.

We formulate our method as a convolutional neural network that predicts a probability distribution over quantities of objects common in geotagged social media imagery from co-located satellite imagery. Though our method relies on social media imagery which is sparsely distributed, it retains the benefits of dense coverage provided by satellite imagery. Our approach can be thought of as learning a deep feature embedding that directly encodes land use without the need to specify a label taxonomy or manually label training data.

II. Approach

In this section, we describe our approach for implicit land use classification by modeling the geospatial distribution of objects visible in ground level images. We follow the cross-view learning framework [3], in which we train a network to understand overhead imagery by having it predict target labels extracted from co-located ground-level images. The result is a method for extracting useful information from overhead imagery without the need for manual annotation.

A. Network Architecture

Our model for predicting the distribution of visible objects from a satellite image is based on the ResNet50 architecture [4]. Given an image as input, we first extract a dense intermediate feature representation using the convolutional layers of this network, resulting in a 2048-D feature. We then introduce two branches, which we refer to as probability distribution heads, one for estimating a distribution of object counts and another for estimating a distribution over image categories.

Each of the probability distribution heads takes as input the intermediate feature representation and is composed of the following layers: two Conv(512, 1)-ReLU layers, followed by a Flatten layer and a Dense(1024)-ReLU layer. Finally, there

is a Dense(N) layer, where N is the number of parameters required to represent the corresponding distribution.

For object counts we model this as 80 independent Poisson distributions. For image categories, the distribution is a Dirichlet over the 1000-simplex. The parameters of each distribution is output by our CNN. During training we minimize the sum of the negative log likelihoods of the resulting distributions.

B. Implementation Details

Our model is trained end-to-end for 24 epochs with a batch size of 64 on an NVIDIA RTX 2080 GPU. We optimize using Adam, clipping gradients to a norm of 1.0. The learning rate cycles linearly between $1e^{-7}$ and $1e^{-5}$ [5] with a period of 8 epochs. We initialize the ResNet50 portion of the network with weights pretrained for the ImageNet classification task [6].

III. Evaluation

We demonstrate that our method is able to capture land use without manually specified categories. Further, we show that these maps are closely related to land use generated by traditional approaches.

A. Dataset

To support our proposed methods, we construct a dataset by extending the recently introduced BigEarthNet [7], which contains Sentinel-2 satellite imagery over ten European countries. For each satellite chip, we identify ground-level images from the Yahoo 100 Million (YFCC100M) [8] dataset whose geotags lie within the chip's bounding box. Each satellite chip typically contains multiple examples from YFCC100M. Due to overlaps in image extents, examples may also lie within multiple satellite chips. From each ground-level image, we extract histograms of object frequency and probability distributions over image categories.

1) Object Counts: We construct histograms representing the frequency of common objects using an existing model pretrained for the MS-COCO object detection task [9]. We apply the model to each geo-tagged image in our dataset, then filter the predictions to those which have a confidence of 0.9 or greater. Finally, we count the number of occurrences of each object category, producing an 80-D vector of object counts.

2) Image Categories: We estimate probability distributions over image categories using an existing model pretrained for the ImageNet classification task [6]. We apply the model to each image in our dataset and perform some moderate temperature scaling [10], scaling the predicted 1000-D logits by a factor of 0.8, to dampen the effect of overly confident predictions.

B. Implicit Land Use Mapping

To qualitatively demonstrate the descriptive power of our learned representation, we construct maps of the learned representation by applying our model to each of the Sentinel-2 satellite images in BigEarthNet. Next, in order to reduce the dimensionality of the predicted representations, we perform principle component analysis (PCA) which identifies a set of orthogonal axes that maximize covariance. In Figure 1 (right), we display a series of maps where the color of each pixel is determined by the value of the first three PCA components. We observe that the generated map is highly correlated with the land use labels defined by CORINE [1]. For example, the red channel appears to be indicative of water sources (oceans, lakes), while the green channel corresponds to forested/wooded areas. Specific geographic features are visible, such as in the first row, the Wicklow Mountains National Park, located south of Dublin, Ireland, can be identified as a noticeably darker patch of the false color map.

TABLE I: Mean classification statistics for the CORINE [1] land use categorization task. Left column: results for training 43 one-versus-all logistic regression classifiers using our learned feature representation. Right column: results for the same approach using features extracted from a ResNet50 trained for the ImageNet task.

Metric (mean)	Ours	ImageNet Features
Accuracy	0.9503	0.9374
F1	0.3344	0.0917
Precision	0.3677	0.2443
Recall	0.3151	0.0665

C. Classifying Land Use

To demonstrate the discriminative power of our approach, we consider the task of learning to classify a taxonomy of land use categories using only our learned feature representation as input. Similar to our proposed training approach, we focus on the BigEarthNet dataset and predict CORINE [1] land use categorization for each image. Since BigEarthNet provides multiple CORINE categories per image, we train 43 one-versus-all logistic regression models. As a baseline, we compare against logistic regression models trained on features extracted from a ResNet50 CNN optimized for ImageNet. In Table I we show that our approach is learning to extract features from satellite imagery that are useful for this task compared to the baseline. BigEarthNet features high class imbalance resulting in similar macro-accuracy scores, thus to provide clarity, we also report F1, Precision, and Recall scores. In Figure 1 (middle), we show maps of predictions alongside ground truth maps of land use.

IV. Conclusions

We proposed a novel approach for implicit land use classification that removes the need for a manually defined label taxonomy. Instead, by focusing on the visual differences evidenced by ground-level imagery we are able to learn features that are informative of land use. Our method operates on satellite imagery and outputs distributions over object counts and image categories. We demonstrated that our method is implicitly capturing the notion of land use.

Acknowledgements: We gratefully acknowledge the financial support of NSF CAREER grant IIS-1553116.

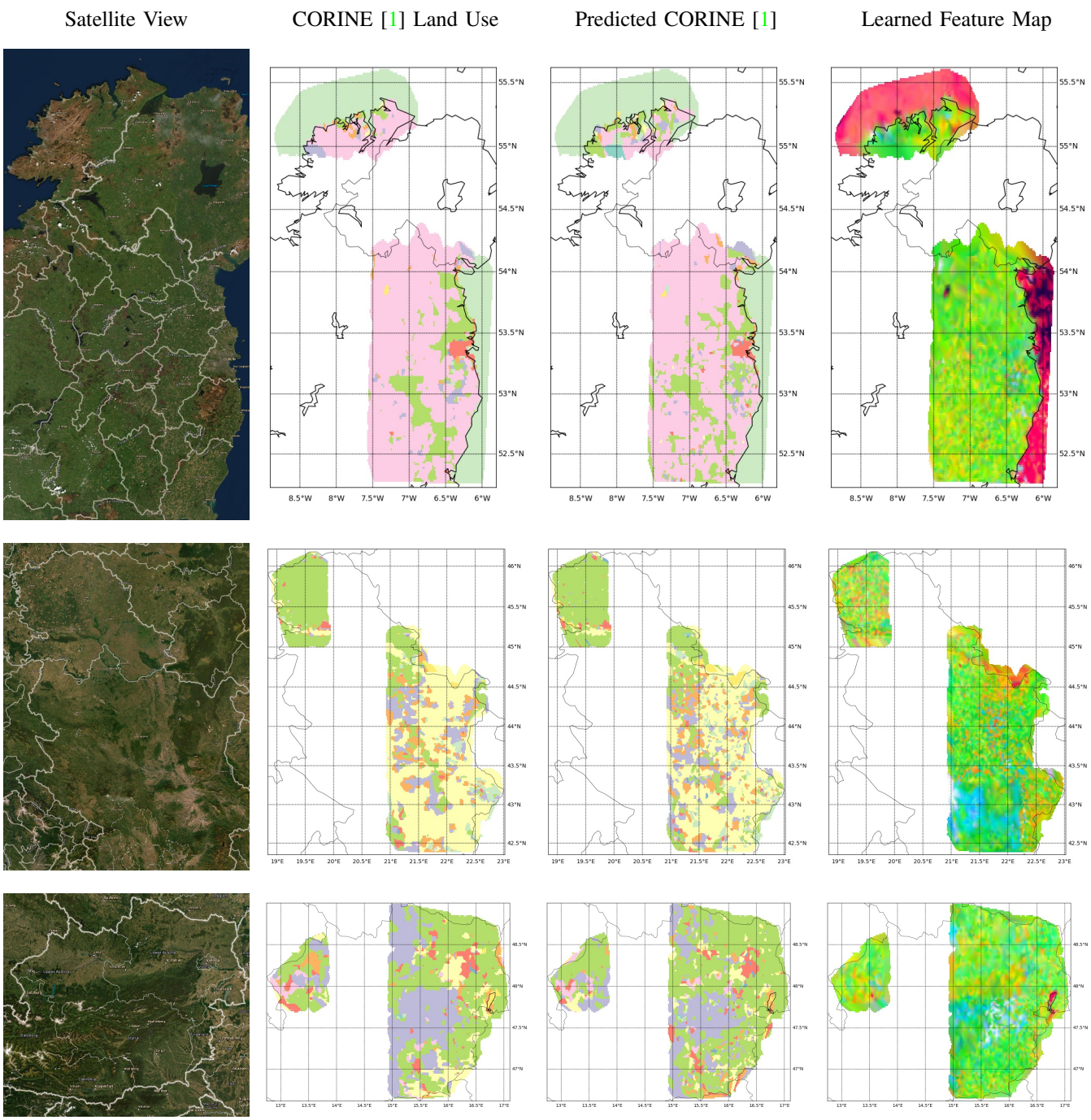

Fig. 1: First column: Sentinel-2 image of the area. Second column: Ground truth CORINE [1] land use category. Third column: Land use category predicted by our model. Fourth column: False-color maps visualizing learned feature representations for the ten geographic regions of the BigEarthNet dataset. (Figure best viewed in color.)

REFERENCES

[1] "CORINE land cover," https://land.copernicus.eu/pan-european/corine-land-cover, accessed: 2019-Nov-9. 1, 2, 3

[2] L. Yang, S. Jin, P. Danielson, C. Homer, L. Gass, S. M. Bender, A. Case, C. Costello, J. Dewitz, J. Fry *et al.*, "A new generation of the united states national land cover database: Requirements, research priorities, design, and implementation strategies," *ISPRS*, vol. 146, pp. 108–123, 2018. 1

[3] S. Workman, R. Souvenir, and N. Jacobs, "Wide-area image geolocalization with aerial reference imagery," in *ICCV*, 2015. 1

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016. 1

[5] L. N. Smith, "Cyclical learning rates for training neural networks," in *WACV*, 2017. 2

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009. 2

[7] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *IGARSS*, 2019. 2

[8] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "YFCC100M: The new data in multimedia research," *arXiv preprint arXiv:1503.01817*, 2015. 2

[9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014. 2

[10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *JMLR*, 2017. 2