Supplementary Information - Large Scale High-Resolution Land Cover Mapping with Multi-Resolution Data

Caleb Robinson
Georgia Institute of Technology

Le Hou
Stony Brook University

Kolya Malkin
Yale University

Rachel Soobitsky
Chesapeake Conservancy

Jacob Czawlytko
Chesapeake Conservancy

Bistra Dilkina
University of Southern California

Nebojsa Jojic
Microsoft Research

April 5, 2019

1 Web application

Here we provide a description and screenshots of the web application described in the main paper.

Users are able to "paint" over existing basemaps (in this case the ESRI World Imagery or OpenStreetMap) with either precomputed model output or on-the-fly model output computed on a server. Users can change the opacity of the painted land cover layer to see how it aligns with the observed imagery, change class weights to create a better fit, and submit the corrections to be included in future rounds of model training or loaded by other users. See `https://aka.ms/cvprlandcover`.

The following figures show different aspects of the application:

- Fig. 1 shows a screenshot of the main window.
- Fig. 2 shows some of the display controls.
- Fig. 3 shows the adaptation and feedback capabilities.

2 Correlation between NLCD and Land Cover classes

In Tab. 1, we show within one NLCD pixel (30 m resolution), the expected distribution of 1 m Land Cover classes, computed using NLCD and Land Cover labels in the state of Maryland.

3 Extended discussion of Consistency maps

In the "National evaluation with NLCD labels" section from the main text we propose a measure, *consistency with NLCD*, for benchmarking high-resolution class predictions in areas for which we only know low-resolution labels (i.e. in areas for which we do not have ground truth high-resolution labels). This type of measure is motivated by the need to visualize how well our model is doing across

	Land Cover Classes			
NLCD Classes	**Water**	**Forest**	**Field**	**Impervious**
Open water	**0.97**	0.01	0.01	0.02
Developed, Open Space	0.00	0.42	**0.46**	0.11
Developed, Low Intensity	0.01	0.31	0.34	**0.35**
Developed, Medium Intensity	0.01	0.14	0.21	**0.63**
Developed, High Intensity	0.01	0.03	0.07	**0.89**
Barren land (Rock/Sand/Clay)	0.09	0.13	**0.45**	0.32
Deciduous Forest	0.00	**0.92**	0.06	0.01
Evergreen Forest	0.00	**0.94**	0.05	0.01
Mixed Forest	0.00	**0.92**	0.06	0.02
Shrub/Scrub	0.00	**0.71**	0.26	0.03
Woody Wetlands	0.01	**0.90**	0.08	0.00
Grassland/Herbaceous	0.01	0.38	**0.54**	0.07
Pasture/Hay	0.00	0.11	**0.86**	0.03
Cultivated Crops	0.00	0.11	**0.86**	0.03
Emergent Herbaceous Wetlands	0.11	0.07	**0.81**	0.01

Table 1: The expected distribution of Land Cover classes (1m resolution) within one NLCD pixel (30m resolution), computed using NLCD and Land Cover labels in the state of Maryland.

the entire US – a scale where simply visualizing and inspecting (input image, output class predictions) tuples becomes infeasible. In this section we elaborate on the consequences of our definition of *consistency with NLCD*, give an alternate definition, and compare the resultant maps from using both definitions on our full-USA predictions.

In the main text, we described our measure as a "charitable" version of accuracy. Indeed, every NLCD class defines a distribution over the high-resolution labels (see Tab. 1) in the Chesapeake bay area, and a given high resolution prediction, y, for NLCD class c, is only considered completely incorrect if $\mu_{c,y} = 0$. This is rarely the case, and could lead to scenarios in which our measure is too generous, e.g. if there is a 30×30 block of all 'Forest' high resolution predictions with a 'Grassland/Herbaceous' NLCD label, then the *consistency with NLCD* value will be $\frac{0.38}{0.54} = 70\%$ which is the ratio of $\mu_{\text{Grassland,Forest}} / \max_{y'} \mu_{\text{Grassland},y'}$ (according to Tbl. 1).

An alternate definition (as briefly mentioned in the main text) is one that more heavily emphasizes deviations from the observed NLCD high-resolution distributions. Here, $\lambda = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{y_i = \max_{y'} \mu_{c_i,y'}\}$. In this definition a prediction is penalized if it does not match the most often observed high resolution label from the associated NLCD label distribution. In the previous example with a block of 'Forest' predictions, this definition of *consistency with NLCD* would give a score of 0.

In Fig. 4 we show maps created with both definitions of our *consistency with NLCD* measure. The areas that show poor performance on both maps should be inspected, as it could be the case that our model is performing poorly there or that the distributions shown in Tbl. 1 do not hold in those areas. One such region in our results is the Great Plains, where our model seems to severely overestimate the 'Forest' class in farmlands. Knowing where and how serious the flaws in our models are is valuable feedback both for model development purposes and to indicate areas that might benefit from manual labeling efforts.

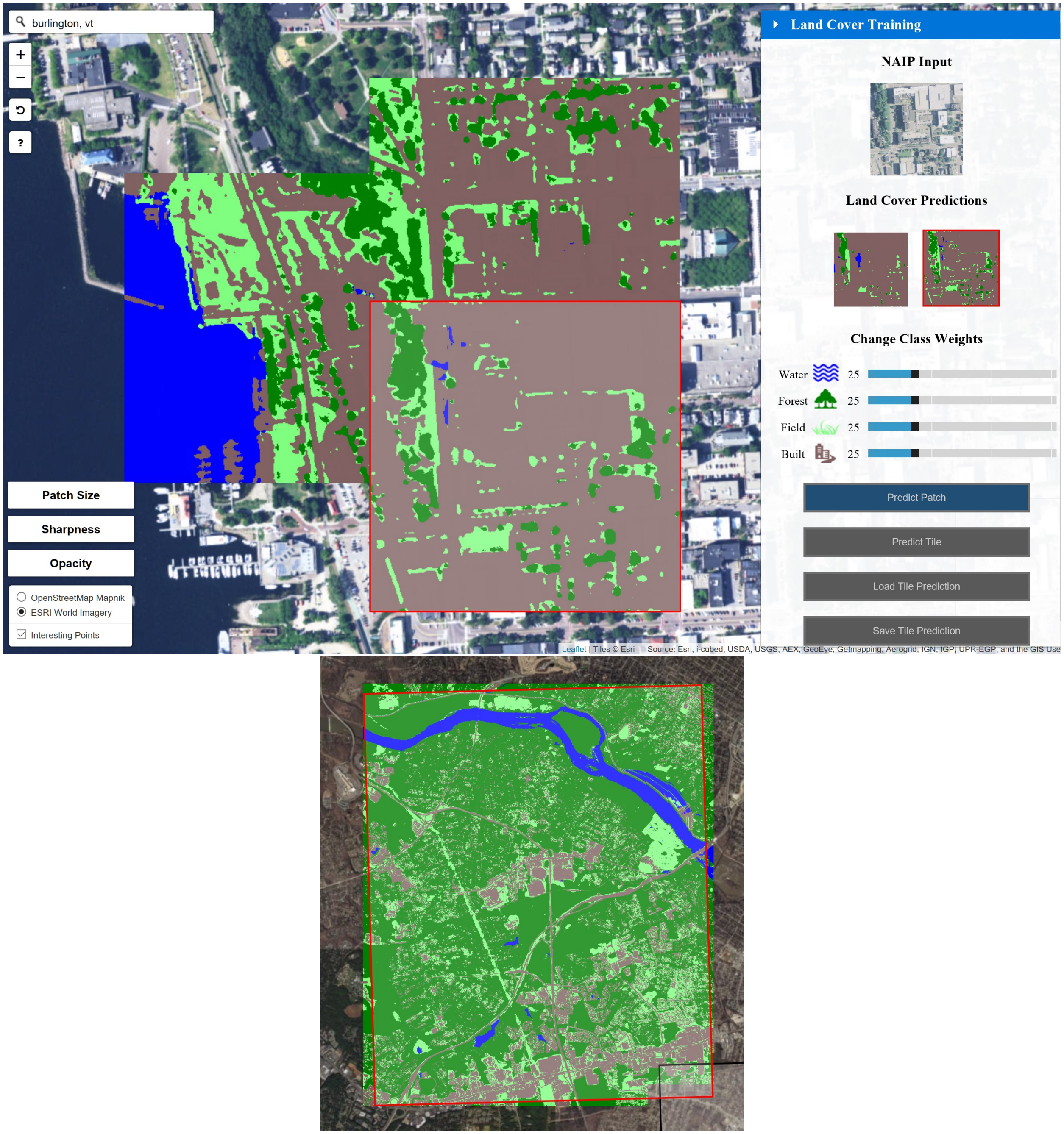


Figure 1: Screenshots of the interactive web application that we have created for interfacing with our model results. Here, a user has used the search box to find a location (Burlington, Vermont) and clicked in three squares, returning land cover predictions (**above**) or loaded predictions for a whole tile (**below**). The functionality of the controls in the right panel and at the lower left corner is demonstrated in Fig. 2.

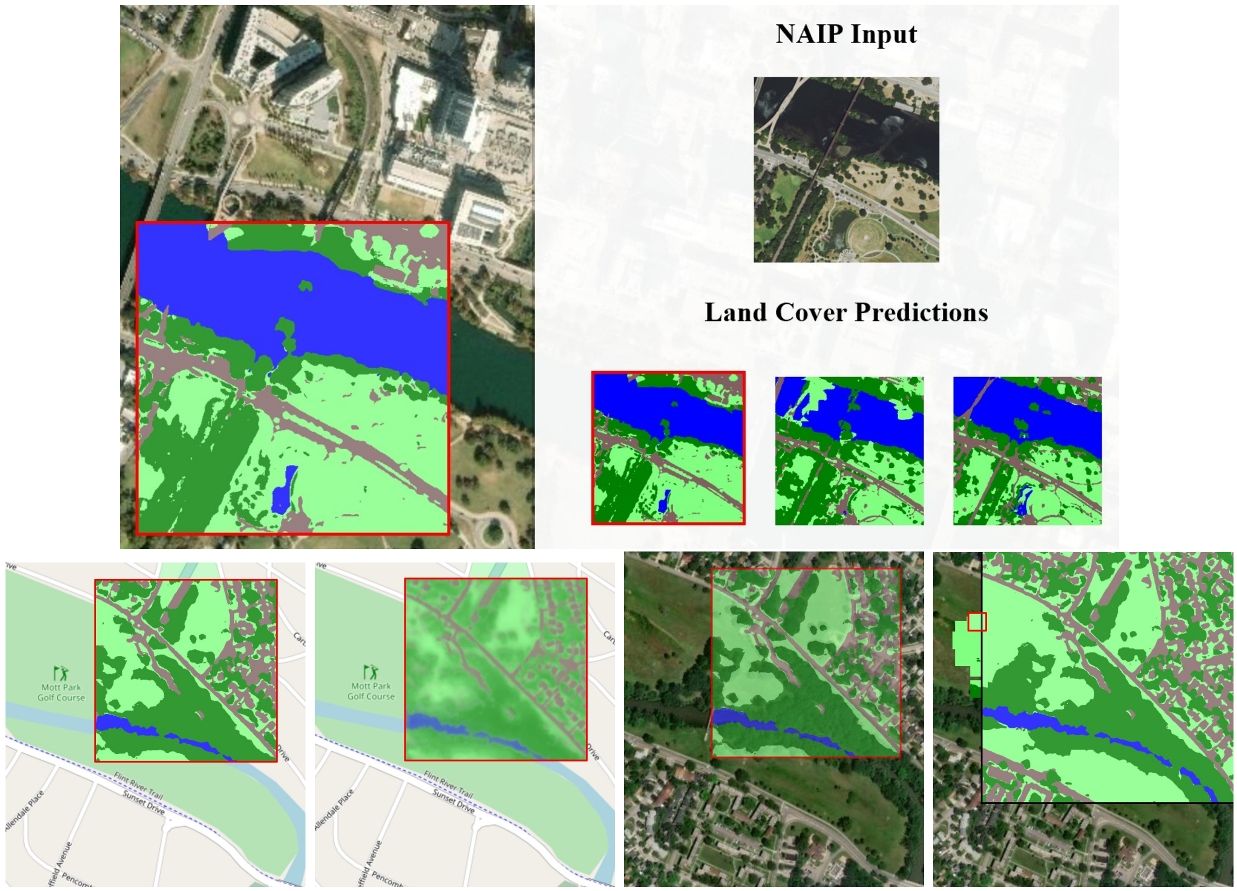


Figure 2: **Above:** By clicking one of three images at right, the user may switch between predictions of two neural models, computed on the fly, and of a third model, loaded from our precomputed map of the entire US. **Below:** Some web app controls, illustrating: selecting argmax or softmax neural net outputs, changing between road and aerial basemaps, modifying the opacity of the predictions drawn over the basemap, and the customizable patch size.

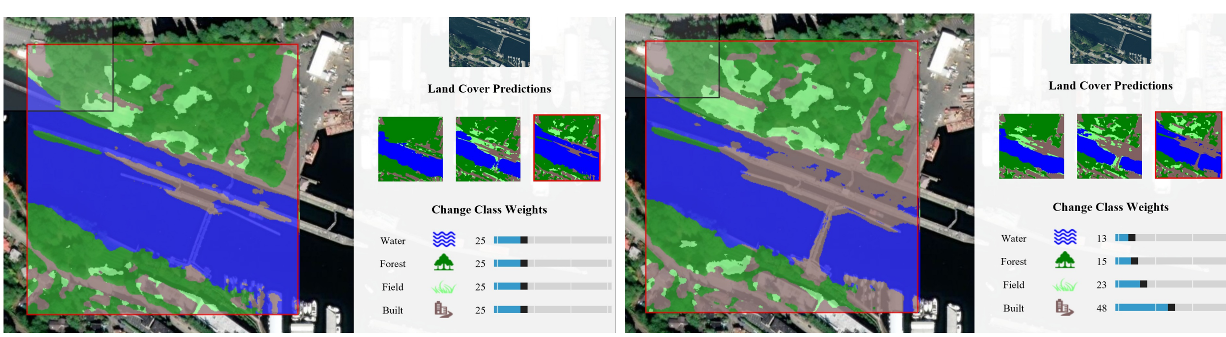


Figure 3: The sliders in the right panel allow the user to weight the neural model's probabilistic outputs for each class. The best predictions may be saved on the server to be used for model adaptation and improvement.

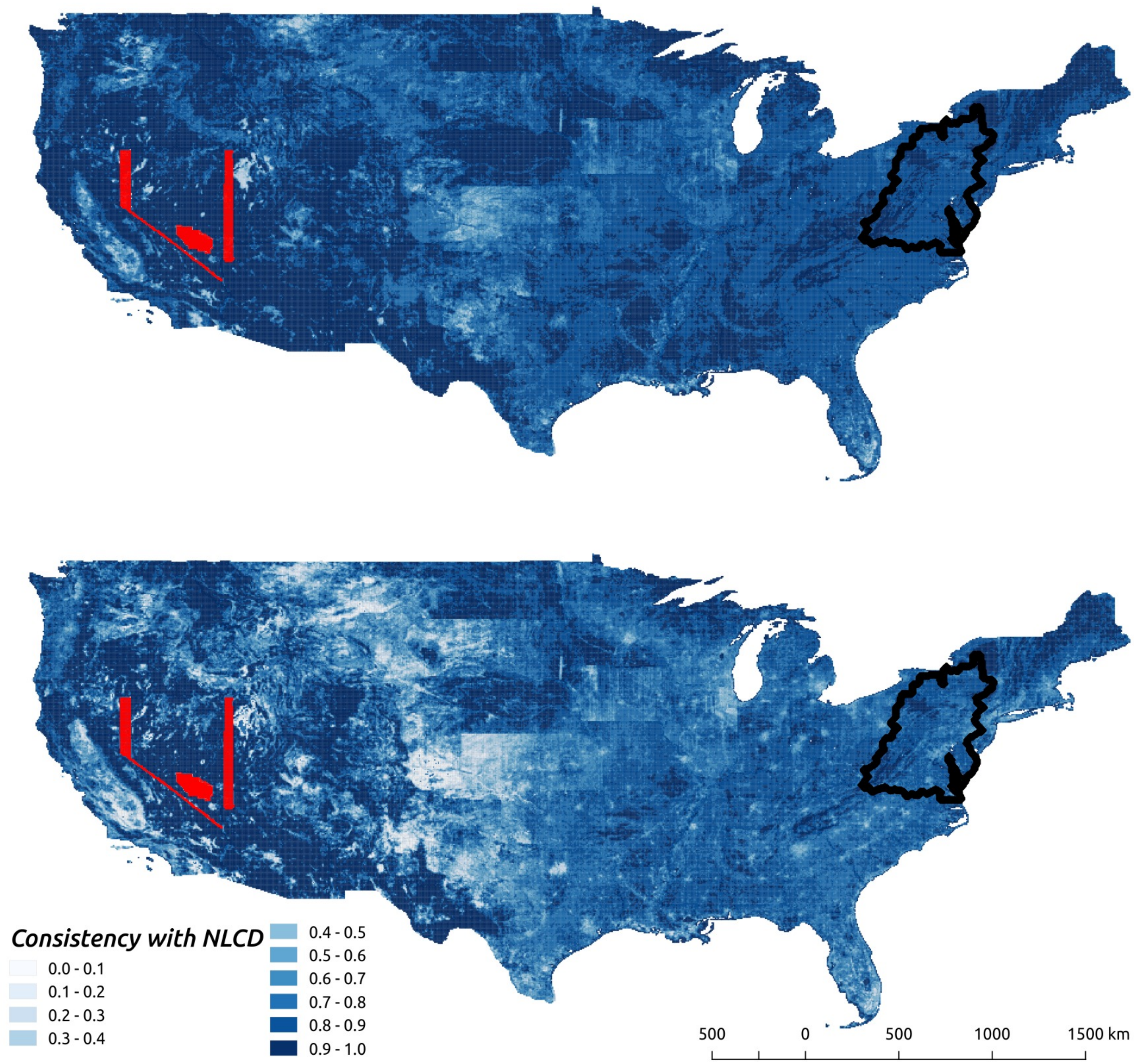
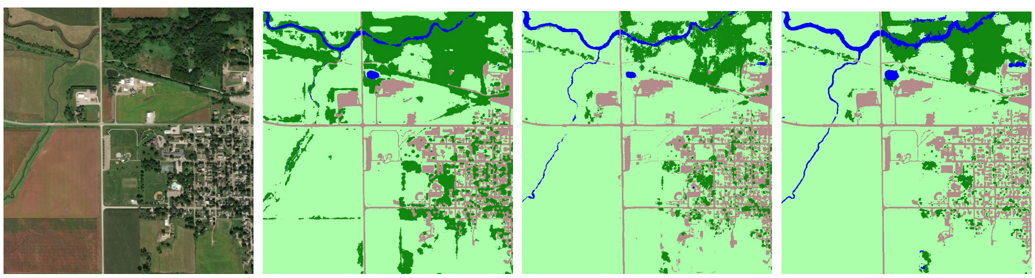


Figure 4: **Lower map** shows the alternate definition of *consistency with NLCD*, while **upper map** shows the *consistency with NLCD* definition given in the main text.



(a) NAIP image from Middle Cedar Watershed in Iowa

(b) Initial automatic land cover prediction

(c) Corrected prediction by the QA process

(d) After re-weighting each class in the initial prediction

Figure 5: Qualitative evaluation of the land cover Quality Assurance (QA) process. Given an initial land cover prediction result, shown in (b), our collaborators used their QA methodology to correct the prediction, shown in (c). Finally, we simply re-weighted the softmax outputs of our model to match the correct predictions as much as possible. This process is very time efficient (no model training). We show re-weighted results in (d).