Automated extraction of map symbology from nineteenth century topographical maps by convolutional neural networks for understanding long-term changes in the extent and distribution of wetlands

Rob O'Hara [a,*], Richa Marwaha [b], Jesko Zimmerman [a], Matthew Saunders [b], Stuart Green [a]

[a] Teagasc Food Research Centre, Spatial Analysis Unit, Ireland, 1st Author - Rob.OHara@teagasc.ie, 3rd Author - Jesko.Zimmermann@teagasc.ie, 5th Author - Stuart.Green@teagasc.ie

[b] Trinity College Dublin, Department of Botany, Ireland, 2nd Author - Marwahar@tcd.ie, 4th Author - SAUNDEM@tcd.ie

* Corresponding author

Keywords: wetlands, u-net, segmentation, historic maps, drainage

Abstract:

Digitised historical maps are widely used in qualitative historical research. On-going advances in image segmentation is resulting in increased exploration of historical maps for quantitative purposes in Earth and natural sciences.

Historical maps contain an abundance of information, which makes them invaluable primary sources for understanding past land use. For example, the detailed topographical and cadastral maps undertaken by the Ordnance Survey Ireland (OSi) from the mid-nineteenth to the early twentieth centuries depict several specific cartographic symbols that indicate historic wetland cover at field level. The quality of these maps could be poor because of image compression and scanning process which makes it difficult to extract the cartographic symbols.

Our pilot study, part of a wider project looking at EO mapping of grassland histories in the Republic of Ireland, explores how semantic segmentation using a fully convolutional network (CNN) can extract wetland symbols into national database of Ireland known as PRIME 2 developed by OSi. Our segmentation workflow attempts to define the extent and distribution of former wetlands in order to determine current field drainage status over a range of organic and mineral soils..

In this study, a U-net CNN architecture was used to automate the extraction of wetland symbology from digitised historical maps. Segmentation used open-source algorithms from the 'pytorch' library in Python. Separate models were trained for Ordnance Survey Ireland digital historic maps: the 1st edition 6-inch colour map (c. 1840s) and the 3rd edition 6-inch black and white map (c. 1940s). For the 1st edition map, an area based mask was used to mark wetlands regions. For the 3rd edition map, individual wetland symbols were masked for four symbol types: osiers (willows), rough pasture, marsh, and whins (gorse). A small number of training images for each type of symbol were extracted from the original digital maps. A simple raster graphics editor (e.g. MS Paint) was used to mask extraneous map objects, symbols or text from the training images. Model parameters were trained using a Stochastic Gradient Descent (SGD) optimiser on a simple training/testing split. For the 1st edition map an additional exponential learning rate scheduler was applied (gamma = 0.79). The 1st edition model stabilised at 25 epochs with a validation loss of 0.08. The models for 3rd edition map stabilised after 100 epochs with the validation loss ranging from 0.02 to 0.04. The trained models were then applied to the maps sheets using a patched approach, where each map sheet was divided into subsets matching the dimensions of the model training data (512x512 for the 1st edition map, and 256x256 for the 3rd edition maps).

The segmented output images were intersected with the Ordnance Survey Ireland Prime 2 geospatial database to get the thematic accuracy. Prime 2 is an object-oriented vector mapping data model that seamlessly covers the Rep. of Ireland, which includes the boundaries of all agricultural fields. Prime 2 polygons were assigned a binary class label (wetland/no wetland) based on whether wetland symbols were present or not within the modern field boundary on either the first or third edition maps. Overall thematic accuracy for first edition map was 80.92% (kappa coefficient of 0.61). Accuracy for the third edition map was 81.28% (kappa coefficient of 0.62). There was a 13% change in the polygon labels from wet to dry between first and 3rd editions. However, 10% of polygons changed from dry in the first edition to wet in the third edition. Possible reasons for this are discussed, as well as common sources of confusion in the segmentation process. Possible improvements to the segmentation workflow are discussed, including pre- and post-processing steps, as well as manual data augmentation to mitigate and reduce instances of misclassification of other map objects and symbols. These

are preliminary results to assess the opportunities of symbol extraction from historical maps for improved understanding of past land use change.
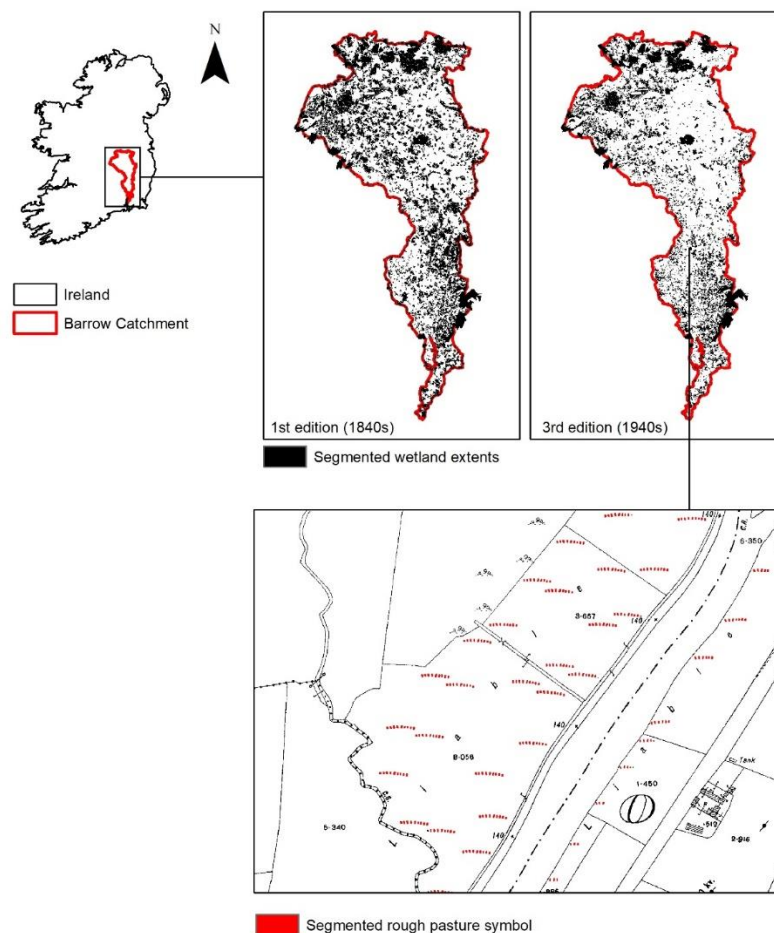

Figure 1. Segmented image from the 3rd edition map.

Work continues on refining the segmentation workflow to reduce the number the false positives and false negatives. Understanding historic land use and management changes on agricultural grasslands, is key to understanding the climate impact these changes are having on agricultural emissions, in particular carbon fluxes from drained organic or organo-mineral soils. These historical data will eventually be combined with optical and RADAR Earth observation time-series for object-based, machine and deep learning classifications of indicative field drainage and management intensity.

Acknowledgements