International Cartographic Association
Association Cartographique Internationale

Improving map text detection and recognition through data synthesis

Samantha T. Arundel[a, *], Trenton P. Morgan[b], Dennis P. Powon[b]

[a] *U.S. Geological Survey, Center of Excellence for Geospatial Information Science, sarundel@usgs.gov*
[b] *Contractor to the U.S. Geological Survey, dpowon@contractor.usgs.gov, tpmorgan@contractor.usgs.gov*

* Corresponding author

Keywords: Deep learning, optical character recognition, historical topographic map, training data, data synthesis

Abstract:

Historical topographic maps present a wealth of information that was collected through rigorous, manual processes and compiled, drawn, and edited by hand. During the digital revolution, mapping agencies began searching for ways to automate the production, and particularly the update, of topographic maps at national scales. Whereas the digital capture and representation of some map elements such as contours lines, structures and roads are fairly straightforward, the opposite is true for features with vague, changing, indeterminate or controversial locations and boundaries (Arundel et al. 2020). As a result, many such features have yet to be depicted on modern, digital, topographic maps. Indeed, the location/extent of these features is often determinable only by considering the historical text properties.

Optical character recognition (OCR) offers a suite of approaches to achieve text detection and recognition. Whereas some research has addressed recognition of built features (Li et al. 2020), the work reported here extends previous efforts to apply deep learning OCR methods to the capture of spot elevation values and locations from the U.S. Geological Survey's historical topographic map collection (HTMC) (Arundel et al. 2021). This endeavor resulted in a small training dataset created through tedious manual digitizing, limiting the dataset to 350, and encouraging recognition results (~80 overall accuracy), although insufficient for an operational procedure (Figure 1). Even more importantly, it highlighted challenges to potentially target in future work, such as limiting the detection to spot elevation numbers rather than other text, recognizing the correct value when numbers were superimposed on contours or other features, and ignoring random spacing differences between characters in a single 'word.' These problems can all be addressed by the addition of much more training data. As the manual creation of the training data labels – the bounding boxes around words and characters and their corresponding numeric value - was too laborious to justify enlarging the dataset using this process, this research presents the enhancement of spatial training data through augmentation and synthesis.

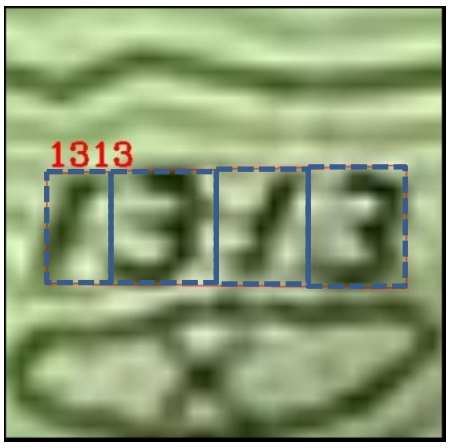

Figure 1. Example of spot elevation image and bounding boxes for training data creation. Bounding boxes were created through tedious manual digitizing, limiting the dataset to 350 image/label pairs.

Abstracts of the International Cartographic Association, 3, 2021.
30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.
https://doi.org/10.5194/ica-abs-3-12-2021

Data augmentation refers to enhancing an existing dataset by image manipulation techniques such as noise reduction, rotation, cropping, flipping, padding, offset, scaling (often referred to as 'cropping') and changes in contrast and brightness. Here we experiment with image rotation, feature offset and scaling (Figure 2). Training images are rotated up to 90 degrees randomly in both directions from center, but no further, as text oriented beyond this is very unlikely to appear on a historical topographic map. Similarly, flipping is not an augmentation option for map text as mirrored text is not equivalent to its original. Feature offset refers to repositioning the target object within the overall image. Always placing the object in the middle of the image frame has been shown to create locational bias in recognition (Li et al. 2021). We experiment with random offsets, which also allow additional non-targeted features to appear in the image and may separately affect the results. Scaling, in this case, refers to varying the map scale at which the image is captured. This technique also varies the appearance of non-target features with corresponding impacts.

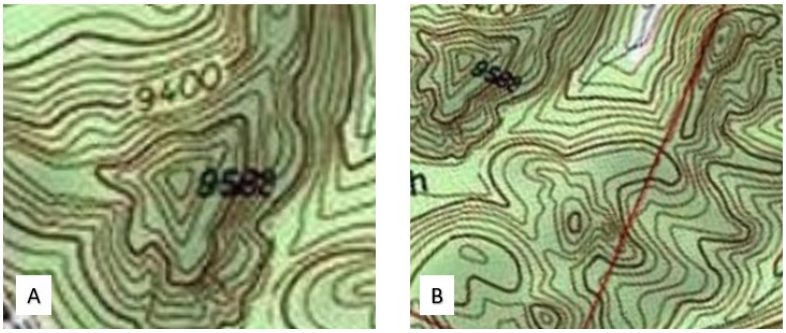

Figure 2. Training images showing (A) original scale, rotation and position and (B) scaling, rotating and offset options.

In data synthesis, images are artificially created rather than being generated by natural phenomena. Synthesis often attempts to digitally (automatically) recreate the original production process. Data synthesis has the advantages of 1) being relatively quick and inexpensive, 2) producing 100% accurate labels, and 3) enabling the use of deep learning approaches to detection in otherwise small training datasets. In this research, we synthesize the images by employing the US Topo (https://www.usgs.gov/core-science-systems/national-geospatial-program/us-topo), which has no spot elevations, as the base image. We then use image science methods to display the corrected GNIS summit point locations and values (Arundel and Sinha 2020) in similar HTMC text font on the base image (Figure *). Some of the challenges to this approach include proper scaling to achieve the desired resolution of the image, determining the best font, color, character size, and resolution to use in order to mimic HTMC text, and automatically creating appropriately sized bounding boxes.

Figure 3. Current best image synthesis output.

The work reported here is in progress, but based on experiences documented in the literature, we expect improved results. Our hope is to achieve an overall accuracy of ~90%, at which point we can begin to understand the necessary workflow to move the research into production.

Abstracts of the International Cartographic Association, 3, 2021.
30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.
https://doi.org/10.5194/ica-abs-3-12-2021 | © Author(s) 2021. CC BY 4.0 License.