MRE AUTOMATED METHODS TO COLLECT HISTORICAL INFORMATION FROM LIDAR IMAGES AND MAPS.

Jeff Blackadar 100139943

HIST 5308 Masters Research Essay

This research essay submitted to Carleton University in fulfillment of the requirements for the course HIST 5908, as credit toward the degree of Master of Arts in History, Specialization in Data Science.

1 March 2022

History is written not just in documents, but in the land - and in documents that *represent* the land. This MRE project presents case studies in using computational methods to surface new historical information about historic landscapes, making a wider variety of historical questions possible. This project is about making new methods and identifying the new kinds of questions these methods permit us to ask.

The case studies are united by a common thread: elements of the landscape that are now largely invisible or forgotten in day-to-day life but have a profound impact on the history of a region. The first case study explores the landscapes of production for the early iron industry in Pennsylvania. The second explores mapmaking in Mandate-era Syria and Lebanon. Finally, preparing these case studies is itself an exercise in replicability and reproducibility in historical research; the MRE's methods sections are in fact tutorials prepared for publication.

The computational methods involved are colloquially known as 'Deep Learning'. Deep Learning technology, a form of Artificial Intelligence, has matured rapidly in recent years. It has reached a point where it can be applied in a practical and reliable manner to recognize a large scope of patterns and information much faster than human beings can. This presents avenues for historical inquiry unavailable to researchers until now. For example, Deep Learning can be used to recognize the remains of human construction present in aerial images such as LiDAR over large areas that would take a human researcher a much longer time to accomplish. Another form of Deep Learning can be used to read text from maps in a systematic way to automate the search for specific types of locations.

Deep learning presents tremendous opportunities for historians, but also challenges. The technology is relatively new and continues to evolve. Documented methods are needed to employ this technology to bring it into practical use, while avoiding pitfalls and being aware of technical limitations. Also, and more importantly, Deep Learning itself presents issues to historians to respond to. Historians have a unique perspective to critique and improve the use of Deep Learning, such as analyzing the training process, examining biases in training data and responding to historical issues with sources of training data, such as colonial maps.

The objective of this MRE is to address those challenges in two ways. I will document the use of Deep Learning for object and text recognition in a manner where other historians can employ the techniques. I will also examine issues to consider when training a Deep Learning model as well as sources. This will fulfil my goals of producing work that is of use to other historians and including methods from Data Science.

The MRE will also have two branches of inquiry based on the examples mentioned above: use of Deep Learning to recognize visible objects in LiDAR images and use of Deep Learning based optical character and symbol recognition to automatically read maps. I have studied both areas as a research assistant.

The focus of my research into object recognition in LiDAR images was to locate the presence of relict charcoal hearths (RCHs) in the landscape of Pennsylvania. RCHs are evidence of the charcoal industry that fueled iron production in the United States from the eighteenth century until the last charcoal burning furnace was shut down in 1945. I have used LiDAR images from Pennsylvania Spatial Data Access (PASDA) to train a Deep Learning model to recognize RCHs in the landscape of that state. I plan to use LiDAR data files from the state of Maryland to compare the efficacy of the Pennsylvania trained model to detect RCHs in a nearby, but different, location.

The research to automatically read historical maps using Deep Learning will use maps for Pennsylvania as well as Syria and Lebanon. For Pennsylvania, Dr. Ben Carter has provided me with historical maps of some charcoal producing regions dating from the mid-nineteenth century. I will document the method used to transcribe these maps using optical character recognition and to see what can be learned from these maps when combined with evidence of known areas of historical charcoal production. I plan to locate similar maps of charcoal producing regions of Maryland and apply the same technique to demonstrate reproducibility of the method and study the production of charcoal in that state.

For my work with maps of Syria and Lebanon, I have access to a set of georeferenced maps of produced by the *Service Géographique des Forces Française Libres du Levant* during 1920s-1950s. These colonial maps are useful to research methods to transcribe French and Arabic text and research the automated detection of symbols. Work with these maps ties into research I have done for Computational Research on the Ancient Near East - Computational Creativity and Archaeological Data project ([CRANE-CCAD](https://shawngraham.github.io/CCAD/)). As an example, I have used the optical character recognition (OCR) method to detect the Arabic spelling of the word *tell,* a mound, which may be archaeologically significant sites of habitation. This is described in the paper “Potential of Deep Learning Segmentation for the Extraction of Archaeological Features from Historical Map Series” by Arnau Garcia‐Molsosa et al.[[1]](#footnote-1) The authors of the paper employ a different technology to read maps and so I hope the method I document will also prove useful.

For Deep Learning with LiDAR, I will draw on work I have done with Dr. Ben Carter and Weston Conner. They are archaeologists and have expertise with the history of charcoal production in Pennsylvania. Using their dataset known RCHs in the landscape of Pennsylvania, as mentioned above, I developed and trained an automated method to detect additional RCHs in Pennsylvania.

Some of this work is already in the publishing pipeline. I will draw on work I have done as a co-author of an article submitted to [Advances in Archaeological Practice](https://www.cambridge.org/core/journals/advances-in-archaeological-practice). Dr. Ben Carter is the lead author, with Weston Conner and I as co-authors. They had the idea of using automated recognition of relict charcoal hearths (RCH) in the landscape of Pennsylvania. They also provided the training data of known locations of RCH in order to train a Deep Learning model. My contribution was to build a method to train a Deep Learning model to locate more RCHs in landscapes across the state and document the method.

Our team also submitted related articles to the [Journal of Open Archaeological Data](https://openarchaeologydata.metajnl.com/). Weston Conner is the lead author on an article describing the LiDAR data set and the methods used to process the data and I am a co-author. I am the lead author on an article describing the dataset and process to build and run the Deep Learning model, including the results. For additional information, I have attached the current versions of the articles.

This MRE will look different than MREs produced as part of the History, Public History program, and the Digital Humanities program. It has a focus on methods and the clear documentation of these methods, for reuse by other historians. I propose to document the methods I have been working with in a manner so that they can be used as tutorials. I will propose to the [Programming Historian](https://programminghistorian.org/) a short series of connected lessons. The raw materials for these lessons are [blog posts](https://jeffblackadar.ca) authored as part of my work as a research assistant for the CRANE-CCAD project. The series of lessons would be as follows:

The first lesson would be instructions about how to connect to a Microsoft service to read handwritten and typed text. This method can be applied to many different types of documents and it will be used as a building block to automatically read maps.

The second lesson will deal with working with georeferenced maps, including the use of coordinate reference systems, splitting maps into smaller tiles and programming considerations. This tiling method can be applied to historical maps as well as modern LiDAR images.

The third lesson would demonstrate the use of OCR to read text from map tiles to produce a file of map place names and their geographic locations. As part of this lesson, I would provide a commentary about issues for a historian to consider when using maps. It is possible the idea for these lessons may not be approved by the Programming Historian. In that case, I would look for another way to publish this, such as a [Gitbook](https://www.gitbook.com/) website.

I would like to acknowledge that the idea to automatically detect RCHs came from Dr. Ben Carter and Weston Conner and without their data and analysis, my work to build a model to detect RCHs would not have happened. Also Dr. Carter inspired my work on the transcription of maps by asking if I could transcribe some historical maps he had. A substantial number of references I have so far come from the work done with Dr. Carter and Weston Conner. I also thank Dr. Kristen Hopper, Dr. Hector Orengo and Dr. Dan Lawrence for providing me the georeferenced maps of Syria and Lebanon produced by the *Service Géographique des Forces Française Libres du Levant.*

Selected References

Primary Sources

LiDAR data files for Pennsylvania. Pennsylvania Spatial Data Access (PASDA) <https://www.pasda.psu.edu/>

Carter, Benjamin and Conner, Weston. Datafile of locations of Relict Charcoal Hearths in Pennsylvania.

Secondary Sources

Books and Journal Articles

Bickler, Simon H. “Machine Learning Arrives in Archaeology.” *Advances in Archaeological Practice* 9, no. 2 (May 2021): 186–91. <https://doi.org/10.1017/aap.2021.6>.

Bonhage, Alexander, Mahmoud Eltaher, Thomas Raab, Michael Breuß, Alexandra Raab, and Anna Schneider. “A Modified Mask Region-Based Convolutional Neural Network Approach for the Automated Detection of Archaeological Sites on High-Resolution Light Detection and Ranging-Derived Digital Elevation Models in the North German Lowland.” *Archaeological Prospection* n/a, no. n/a. Accessed May 30, 2021. <https://doi.org/10.1002/arp.1806>.

Bonhage, Alexander, Florian Hirsch, Thomas Raab, Anna Schneider, Alexandra Raab, and Will Ouimet. “Characteristics of Small Anthropogenic Landforms Resulting from Historical Charcoal Production in Western Connecticut, USA.” *CATENA* 195 (December 1, 2020): 104896. <https://doi.org/10.1016/j.catena.2020.104896>.

Carter, Benjamin P. “Identifying Landscape Modification Using Open Data and Tools: The Charcoal Hearths of the Blue Mountain, Pennsylvania.” *Historical Archaeology* 53, no. 2 (June 1, 2019): 432–43. <https://doi.org/10.1007/s41636-019-00171-1>.

Casana, Jesse. “Regional-Scale Archaeological Remote Sensing in the Age of Big Data: Automated Site Discovery vs. Brute Force Methods.” *Advances in Archaeological Practice* 2, no. 3 (August 2014): 222–33. <https://doi.org/10.7183/2326-3768.2.3.222>.

Conner, Weston. *Industrial Landscapes Lost in the Wild: Creating a Methodology for the Remote Identification of Charcoal Hearths*, 2018. <http://sal.muhlenberg.edu:8080/librarydspace/handle/10718/3428>.

“Crown Land Information Management Centre | Novascotia.Ca.” Accessed May 30, 2021. https://novascotia.ca/natr/land/grantmap.asp.

Davis, Dylan S. “Geographic Disparity in Machine Intelligence Approaches for Archaeological Remote Sensing Research.” *Remote Sensing (Basel, Switzerland)* 12, no. 6 (2020): 921-. <https://doi.org/10.3390/rs12060921>.

Davis, F. A, and H. L Kochersperger. Illustrated Historical Atlas of Berks County, Penna. Reading, PA: 1876.

Dombrowski, Quinn, Tassie Gniady, and David Kloster. “Introduction to Jupyter Notebooks.” *Programming Historian*, December 8, 2019. <https://programminghistorian.org/en/lessons/jupyter-notebooks>.

Fernandez-Diaz, Juan, William Carter, Ramesh Shrestha, and Craig Glennie. “Now You See It… Now You Don’t: Understanding Airborne Mapping LiDAR Collection and Data Product Generation for Archaeological Research in Mesoamerica.” *Remote Sensing* 6 (October 1, 2014): 9951–10001. <https://doi.org/10.3390/rs6109951>.

Garcia‐Molsosa, Arnau, Hector A. Orengo, Dan Lawrence, Graham Philip, Kristen Hopper, and Cameron A. Petrie. “Potential of Deep Learning Segmentation for the Extraction of Archaeological Features from Historical Map Series.” *Archaeological Prospection*, Accessed May 8, 2021. <https://doi.org/10.1002/arp.1807>.

Green, Orengo, Alam, Garcia-Molsosa, Conesa, Ranjan, Singh, and Petrie. “Re-Discovering Ancient Landscapes: Archaeological Survey of Mound Features from Historical Maps in Northwest India and Implications for Investigating the Large-Scale Distribution of Cultural Heritage Sites in South Asia.” *Remote Sensing (Basel, Switzerland)* 11, no. 18 (2019): 2089-. <https://doi.org/10.3390/rs11182089>.

Johnson, Katharine M., and William B. Ouimet. “Rediscovering the Lost Archaeological Landscape of Southern New England Using Airborne Light Detection and Ranging (LiDAR).” *Journal of Archaeological Science* 43 (2014): 9–20. <https://doi.org/10.1016/j.jas.2013.12.004>.

———. “Rediscovering the Lost Archaeological Landscape of Southern New England Using Airborne Light Detection and Ranging (LiDAR).” *Journal of Archaeological Science* 43 (March 2014): 9–20. <https://doi.org/10.1016/j.jas.2013.12.004>.

Kent, Alexander James, Soetkin Vervust, Imre Josef Demhardt, and Nick Millea. *Mapping Empires: Colonial Cartographies of Land and Sea 7th International Symposium of the ICA Commission on the History of Cartography, 2018*. 1st ed. 2020. Publications of the International Cartographic Association (ICA). Cham: Springer International Publishing, 2020. <https://doi.org/10.1007/978-3-030-23447-8>.

Opitz, Rachel, and Jason Herrmann. “Recent Trends and Long-Standing Problems in Archaeological Remote Sensing.” *Journal of Computer Applications in Archaeology* 1, no. 1 (2018): 19–41. <https://doi.org/10.5334/jcaa.11>.

Opp, James. “The Colonial Legacies of the Digital Archive: The Arnold Lupson Photographic Collection.” *Archivaria*, 2008, 3–19.

Orengo, Hector A., Francesc C. Conesa, Arnau Garcia-Molsosa, Agustín Lobo, Adam S. Green, Marco Madella, and Cameron A. Petrie. “Automated Detection of Archaeological Mounds Using Machine-Learning Classification of Multisensor and Multitemporal Satellite Data.” *Proceedings of the National Academy of Sciences* 117, no. 31 (August 4, 2020): 18240–50. <https://doi.org/10.1073/pnas.2005583117>.

Panich, Lee M., Tsim D. Schneider, and R. Scott Byram. “Finding Mid-19th Century Native Settlements: Cartographic and Archaeological Evidence from Central California.” *Journal of Field Archaeology* 43, no. 2 (2018): 152–65. <https://doi.org/10.1080/00934690.2017.1416849>.

Raab, A., A. Bonhage, A. Schneider, T. Raab, H. Rösler, K.-U. Heußner, and F. Hirsch. “Spatial Distribution of Relict Charcoal Hearths in the Former Royal Forest District Tauer (SE Brandenburg, Germany).” *Quaternary International* 511 (2019): 153–65. <https://doi.org/10.1016/j.quaint.2017.07.022>.

Raab, A., M. Takla, T. Raab, A. Nicolay, A. Schneider, H. Rösler, K.-U. Heußner, and E. Bönisch. “Pre-Industrial Charcoal Production in Lower Lusatia (Brandenburg, Germany): Detection and Evaluation of a Large Charcoal-Burning Field by Combining Archaeological Studies, GIS-Based Analyses of Shaded-Relief Maps and Dendrochronological Age Determination.” *Quaternary International* 367 (2015): 111–22. <https://doi.org/10.1016/j.quaint.2014.09.041>.

Risbøl, Ole, Ole Martin Bollandsås, Anneli Nesbakken, Hans Ole Ørka, Erik Næsset, and Terje Gobakken. “Interpreting Cultural Remains in Airborne Laser Scanning Generated Digital Terrain Models: Effects of Size and Shape on Detection Success Rates.” *Journal of Archaeological Science* 40, no. 12 (2013): 4688–4700. <https://doi.org/10.1016/j.jas.2013.07.002>.

Sărășan, Adriana, Adrian-Cristian Ardelean, Andrei Bălărie, Ruben Wehrheim, Kubatbek Tabaldiev, and Kunbolot Akmatov. “Mapping Burial Mounds Based on UAV-Derived Data in the Suusamyr Plateau, Kyrgyzstan.” *Journal of Archaeological Science* 123 (November 1, 2020): 105251. <https://doi.org/10.1016/j.jas.2020.105251>.

Shbita, Basel, Craig A. Knoblock, Weiwei Duan, Yao-Yi Chiang, Johannes H. Uhl, and Stefan Leyk. “Building Linked Spatio-Temporal Data from Vectorized Historical Maps.” In *The Semantic Web*, edited by Andreas Harth, Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Heiko Paulheim, Anisa Rula, Anna Lisa Gentile, Peter Haase, and Michael Cochez, 409–26. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020. <https://doi.org/10.1007/978-3-030-49461-2_24>.

Somrak, Maja, Sašo Džeroski, and Žiga Kokalj. “Learning to Classify Structures in ALS-Derived Visualizations of Ancient Maya Settlements with CNN.” *Remote Sensing* 12, no. 14 (January 2020): 2215. <https://doi.org/10.3390/rs12142215>.

Uhl, Johannes H., Stefan Leyk, Yao-Yi Chiang, Weiwei Duan, and Craig A. Knoblock. “Automated Extraction of Human Settlement Patterns From Historical Topographic Map Series Using Weakly Supervised Convolutional Neural Networks.” *IEEE Access* 8 (2020): 6978–96. <https://doi.org/10.1109/ACCESS.2019.2963213>.

Vermeulen, Frank, Marc Antrop, Beatrijs Hageman, and Torsten Wiedemann. “Ancient Roads and Fields in Northwestern Gaul – A GIS-Based Analysis,” n.d., 10.

Werbrouck, I., M. Antrop, V. Van Eetvelde, C. Stal, Ph De Maeyer, M. Bats, J. Bourgeois, et al. “Digital Elevation Model Generation for Historical Landscape Analysis Based on LiDAR Data, a Case Study in Flanders (Belgium).” *Expert Systems with Applications* 38, no. 7 (2011): 8178–85. <https://doi.org/10.1016/j.eswa.2010.12.162>.

Maps

“1:200,000 Levant. Sheet NI-37-VII., Damas. 1941.” Service Géographique des Forces Française Libres du Levant, 1941.

“Aafrine.” Syria EPSG:4326, 36.71 W, 36.40 S, 37.03 E, 36.78 N: Service Géographique des Forces Française Libres du Levant, 1942.

“Aarsal.” Syria EPSG:4326 36.21 W, 33.94 S, 36.53 E, 34.29 N: Service Géographique des Forces Française Libres du Levant, 1945.

“Amouk.” Syria EPSG:4326 36.20 W, 36.14 S, 36.53 E, 36.53 N: Service Géographique des Forces Française Libres du Levant, 1942.

“Banias.” Syria EPSG:4326 35.69 W, 34.90 S, 36.02 E, 35.28 N: Service Géographique des Forces Française Libres du Levant, 1943.

“Batroun.” Syria EPSG:4326 35.46 W, 34.19 S, 35.77 E, 34.53 N: Service Géographique des Forces Française Libres du Levant, 1943.

“BentJbail.” Syria EPSG:4326 35.21 W, 32.94 S, 35.52 E, 33.29 N: Service Géographique des Forces Française Libres du Levant, 1959.

“Beyrouth.” Lebanon EPSG:4326, 35.41 W, 33.68 S, 35.78 E, 34.04 N: Service Géographique des Forces Française Libres du Levant, 1940.

“Djeble.” Service Géographique des Forces Française Libres du Levant, 1942.

**Selected References**

**Books and Journal Articles**

**Maps**

**Websites**

1. Arnau Garcia‐Molsosa, Hector A. Orengo, Dan Lawrence, Graham Philip, Kristen Hopper, and Cameron A. Petrie, “Potential of Deep Learning Segmentation for the Extraction of Archaeological Features from Historical Map Series,” Archaeological Prospection, Accessed May 8, 2021. <https://doi.org/10.1002/arp.1807>.

   I learned the significance of the word *tell* from this article. [↑](#footnote-ref-1)