A 3D POINT CLOUD DEEP LEARNING APPROACH USING LIDAR TO IDENTIFY ANCIENT MAYA ARCHAEOLOGICAL SITES

Heather Richards-Rissetto[a], David Newton[b], Aziza Al Zadjali[c]

[a] School of Global Integrative Studies (SGIS), University of Nebraska-Lincoln, Nebraska, USA (richards-rissetto@unl.edu)
[b] College of Architecture, University of Nebraska-Lincoln, Nebraska, USA (david.newton@unl.edu)
[c] Department of Computer Science & Engineering, University of Nebraska-Lincoln, Nebraska, USA
(aziza.alzadjali@huskers.unl.edu)

KEY WORDS: Ancient Maya, Archaeology, Deep Learning, LIDAR, Point Clouds, 3D Shape Classification

ABSTRACT:

Airborne light detection and ranging (LIDAR) systems allow archaeologists to capture 3D data of anthropogenic landscapes with a level of precision that permits the identification of archaeological sites in difficult to reach and inaccessible regions. These benefits have come with a deluge of LIDAR data that requires significant and costly manual labor to interpret and analyze. In order to address this challenge, researchers have explored the use of state-of-the-art automated object recognition algorithms from the field of deep learning with success. This previous research, however, has been limited to the exploration of deep learning processes that work with only 2D data, which excludes the use of available 3D data. Our research addresses this gap and contributes knowledge on the use of deep learning-based processes that can classify archaeological sites from LIDAR generated 3D point cloud datasets. LIDAR data from the UNESCO World Heritage Site of Copan, Honduras is used as the primary dataset to compare the classification accuracy of deep learning models using 2D and 3D data. The results demonstrate that models using 3D point cloud datasets provide the greatest classification accuracy in identifying Maya archaeological sites while requiring less data preparation. Further, the research contributes knowledge on the efficacy of data augmentation strategies when working with small 3D datasets.

1. INTRODUCTION

Remote sensing technologies, such as Light Detection and Ranging (LIDAR), are transforming archaeology. Airborne LIDAR Systems (ALS) use a laser pulse emitted from an airplane to collect 3D measurements of landscapes by calculating Time-of-Flight (ToF); that is, the time it takes light to hit objects in the landscape and reflect to the scanner. Unlike satellite imagery, ALS can penetrate canopied forests to reveal the ground surface below the canopies. This allows researchers to detect and map unknown archaeological sites in areas never before possible, and is especially true for surveys in Central America that have focused on ancient Maya landscapes in Guatemala, Honduras, Belize, and Mexico (e.g., Chase and Chase 2017; Prufer and Thompson 2016; von Schwerin et al. 2016). ALS surveys in these areas have produced massive amounts of data that archaeologists must manually process to identify archaeological features. This task is extremely time-consuming, expensive, and requires a high-level of technical and archaeological expertise. Consequently, significant amounts of LIDAR data from this region have not been post-processed. The development of automated methods to identify archaeological features in such datasets is therefore a pressing problem in the field.

Deep learning (DL) is a sub-field of machine learning that has demonstrated state-of-the-art performance on automated object recognition tasks involving both 2D and 3D datasets. DL models use layers of artificial neurons to construct a piecewise function that can iteratively adjust its weights through optimization procedures to map a set of inputs to a set of outputs. The pattern and design of these layers is referred to as the architecture of the DL model. Convolutional Neural Networks (CNNs)—a DL architecture developed specifically for 2D image datasets—have out-performed competing approaches for object recognition. CNNs have been used for the analysis of 2D remote sensing data in a variety of fields for tasks as wide ranging as pollution detection to poverty estimation (Piaggesi et al. 2019, Li et al. 2016). In the field of archaeology, CNNs have been used to automate the identification of archaeological sites from ALS data (Albrecht et al. 2019, Bundzel et al. 2020, Guyot et al. 2021, Kazimi et al. 2018, Somrak, Džeroski, and Kokalj 2020, Verschoof-van der Vaart et al. 2020). This previous research, however, has been limited to DL architectures that employ 2D data. This approach excludes a wealth of available 3D data, such as point cloud datasets, that could aid in the detection and classification of archaeological sites.
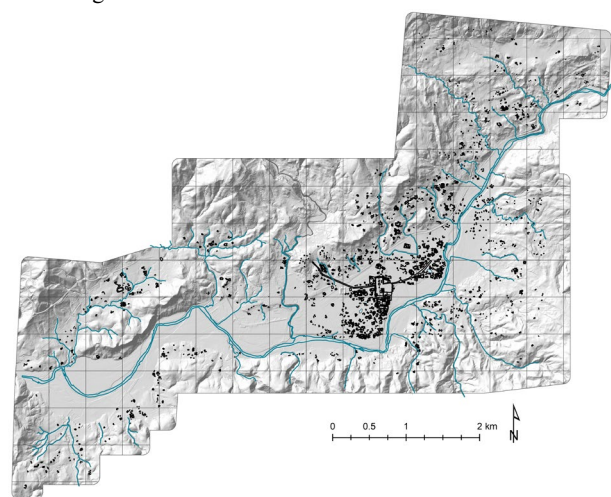

Figure 1. Copan ALS survey area illustrating archaeological sites on LIDAR-derived Digital Elevation Model (DEM).

This research addresses this gap in previous work through the application of DL processes that can classify Maya archaeological sites from ALS generated 3D point cloud

datasets. Specifically, the PointConv (Wu, Qi, and Fuxin 2019) DL architecture is used to identify ancient Maya archaeological sites from ALS-generated 3D point cloud data taken from 26 square kilometers surrounding the UNESCO World Heritage Site of Copan, Honduras (Figure 1). This method is tested against CNN-based processes that rely on 2D data to determine the most efficacious approach. In addition, data augmentation strategies for working with small 3D datasets are evaluated. The results of these experiments demonstrate that the PointConv architecture provides greater classification accuracy in identifying Maya archaeological sites than the CNN-based approach. This result demonstrates a path for researchers to make use of 3D point cloud data directly in DL models while improving accuracy and reducing data preparation time.

The research is organized into the following sections: study area and dataset (Section 2); overview of previous research in deep learning and archaeology (Section 3); description of the deep learning models, experimental set-ups, and data preparation (Section 4); results and discussion (Section 5); lastly, a summary of conclusions and future directions (Section 6).

2. CASE STUDY: MAYA SITE OF COPAN, HONDURAS

2.1 Historical Background

As early as 1000-800 BCE, the ancient Maya constructed massive and elaborate ceremonial architecture. In fact, due to ALS, archaeologists recently located the oldest and largest construction found in the Maya area (Inomata et al. 2020). However, not only is ALS assisting archaeologists in site reconnaissance, but it is also changing our views about the scale of the ancient Maya civilization with the identification of thousands of undocumented sites (Canuto et al. 2018; Garrison et al. 2018). Through LIDAR data along with excavation, epigraphic, iconographic, architectural, and other lines of evidence, archaeologists are developing a clearer picture of the extent and ways the ancient Maya successfully engineered and transformed the landscape for over two millennia.

The case study, Copan, is a UNESCO World Heritage site in Honduras, but from the 5th–9th centuries CE, the city was the cultural and commercial center of a powerful ancient Maya kingdom. The city has awed explorers, archaeologists, and visitors since the 1500s and is the most thoroughly excavated Maya site, providing fundamental data for archaeological interpretations. In 427 CE, *Yax Kuk Mo*, became Copan's first dynastic ruler founding a dynasty that encompassed sixteen rulers and spanned almost four-hundred years (Fash 2001).

2.2 Environment

A key impetus for Maya expansion into the Copan area was its favourable environment with an abundance of rich, natural resources including highly productive soils and sufficient rainfall for agriculture, localized clay sources for ceramics, and plentiful wildlife as well as its location along a critical trade route. The city of Copan is located along the Copan River in the elongated Copan Pocket with an approximated average width of four kilometers. Within this circumscribed environment, altitude ranges from 569-1408 meters resulting in varied topography (flat to steep), diverse vegetation and ecological zones, and differential landuse practices—in the past and today. This ecologically and topographically diverse landscape leads to additional challenges in post processing LIDAR data.

2.3 Project Data: Archaeological and Lidar

In the late 1970s and early 1980s, the Copan Archaeological Project carried out a "100%" pedestrian survey of the Copan Pocket and instrument-mapped almost 600 sites within twenty-four square kilometers surrounding the city's main civic-ceremonial core. This work resulted in a (paper) publication that included 24 one-square kilometer maps that plot archaeological sites, contour lines, and hydrology at a scale 1:2000 (Fash and Long 1983). These paper maps were scanned, georeferenced, and digitized to create a shapefile with attributes of Copan's archaeological structures (n~3500) (Richards-Rissetto 2010).

In 2013, the MayaArch3D Project commissioned the acquisition of LIDAR data (LAS and ASCII) for the Copan Pocket (von Schwerin et al. 2016). Watershed Sciences Inc. (WSI) from Oregon, USA acquired the data using a Leica ALS50 Phase II system mounted in a Piper Aztec aircraft. The target point density was $\geq$ 15 pulses/m^2 and all areas were surveyed with an opposing flight line sidelap overlap of $\geq$ 50%. The average first-return density for the LIDAR data was 21.57 points/m^2 and ground return density averaged 2.91 points/m^2. WSI set two permanent survey monuments and used a Trimble R7 base unit and a roving Trimble R8 GNSS receiver to collect GPS data across the valley. Following acquisition, this LIDAR data went through several (time-consuming) stages of post-processing that incorporated "standard" bare-earth algorithms, semi-automatic, and manual methods to classify 3D points into four classes: (1) Vegetation (green), (2) Default/unclassified (green) , (3) Ground (yellow), (4) Archaeological Features (red), and (5) Ruin Grounds (purple) (see von Schwerin et al 2016 for additional details on post-processing) (Figure 2). The classified data overlaid with the structure shapefile served as the 3D training set in this research (Figure 2). The resultant classified 3D data were used to generate a (2.5D raster) Digital Elevation Model incorporating bare-earth and archaeological mounds that was converted into a hillshade, exported as a TIFF for the 2D (image) training set.
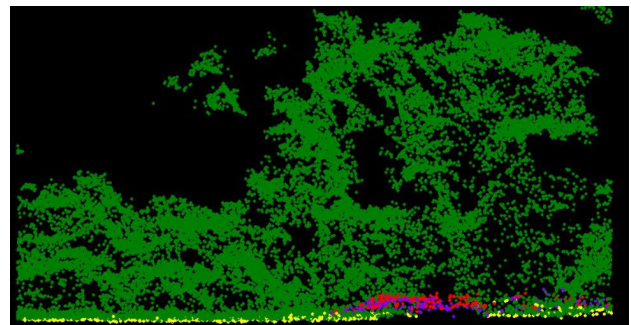

Figure 2. Copan ALS survey 3D Classified Point Clouds.

3. PREVIOUS WORK

3.1 Deep Learning: Object Classification and Semantic Segmentation

The field of computer vision has studied the problem of automating object classification and semantic segmentation for decades (Khurana et al. 2016). Object classification refers to the automated classification of objects in 2D and 3D data. Semantic segmentation, in contrast, refers to the process of classifying all pixels, or points, within a 2D image, or 3D, as belonging to one object contained in a scene verses other objects in that same scene. CNNs were first introduced by LeCun et al. (1989) and

have emerged as the leading approach for object classification and semantic segmentation when working with 2D images (Krizhevsky, Sutskever, and Hinton 2012; Zhao et al. 2017).

A challenge in working with CNNs is that they require large amounts of labeled training data often in the range of millions of images that are pre-labeled and/or segmented by hand. This poses a difficulty when working with small datasets. In order to address this problem, researchers have developed a technique called "transfer learning" (Yosinski et al. 2014). Transfer learning involves repurposing CNN models trained on one image dataset for another image dataset. This is done by retraining only a few selected layers of the CNN model on the new dataset. Transfer learning has demonstrated the ability to improve model accuracy in situations where the available data to train CNN models is small (i.e., thousands of images) (Yosinski et al. 2014). Remote sensing datasets are often small and previous research suggests that the application of transfer learning in this context improves model accuracy (Pires de Lima and Marfurt 2020).

There are also DL architectures specifically designed to work with 3D remote sensing data, such as point clouds. Previous work in this area has explored a variety of DL architectures to perform 3D shape classification and 3D point cloud segmentation tasks (Guo et al. 2020). In terms of 3D shape classification, previous research can be organized into three categories: multi-view methods - which use a combination of 3D point data and 2D image views of objects; volumetric-based methods – which use voxels to represent 3D objects; and point-based methods. Of these, point-based methods have demonstrated some of the highest accuracy when working with 3D point cloud data. Our research, therefore, uses a point-based DL architecture to classify 3D point cloud data in the identification of archaeological sites.

3.2 Deep Learning Archaeological LIDAR Applications

DL is still a relatively new field and its applications in archaeology to identify archaeological sites from remote sensing data, such as LIDAR data, have been limited to the use of DL architectures that work with 2D data. This previous research can be organized into two different categories: approaches using DL architectures for object classification and those that focus on semantic segmentation. The first category provides a probability as an output as to whether an input image is likely to contain an archaeological site. The second approach provides pixel-level probabilities to classify each pixel of the image as belonging to an archaeological structure or not. This method, therefore, provides the classification and the position of archaeological structures within the larger input image. Both approaches can also provide probabilities that can be used to classify the type of archaeological structures (e.g., mound, platform, building, etc.) found in an input image.

In terms of approaches that investigate DL architectures for object classification, Kazimi et al. (2018) test three different CNN architectures for the classification of four different landscape features in the Harz mining Region in Lower Saxony using 2D height maps extracted from LIDAR data. They find that the simplest CNN model, which did not use transfer learning, was more accurate than the VGG-16 and autoencoder architectures. Somrak, Džeroski, and Kokalj (2020) test variations of the VGG-19 CNN architecture with transfer learning to classify LIDAR data into four different categories of Maya archaeological structures. The research also explores

which 2D image formats extracted from LIDAR data improve model accuracy the most. They find that visualization for archaeological topography (VAT) images without analytical hillshading provide the best prediction performance.

Previous research has also explored the use of semantic segmentation to identify archaeological sites. Verschoof-van der Vaart and Lambers (2019) demonstrate the use of the Faster R-CNN semantic segmentation architecture to identify different classes of archaeological structures in the Netherlands from Simple Local Relief Model (SLRM) images derived from LIDAR data. The research provides insight into how small dataset sizes can pose challenges for training DL models. Bundzel et al. (2020) tests two different semantic segmentation DL architectures, U-Net and Mask R-CNN, to identify Maya archaeological sites of various sizes from digital elevation model (DEM) images and find that the U-Net architecture, which does not use transfer learning, preforms best on all sizes of archaeological structures.

This previous research touches on several key issues in the application of DL to identify archaeological structures in LIDAR data. The first issue is the impact that dataset sizes have on the accuracy of DL models. The LIDAR datasets used in this previous research were all relatively small (i.e., in the hundreds and thousands of data samples) compared to the dataset sizes used to get peak accuracy in the training of state-of-the-art CNN models (i.e., in the millions of images). To address this issue, previous research investigated the impact of simplifying DL architectures, so they have less trainable parameters; using transfer learning; and/or applying data augmentation strategies. Data augmentation is a process in which additional training samples are created by copying and modifying (e.g., through scaling, rotation, translation, adding noise, zooming, etc.) existing data to produce new data instances.

Another important issue involves the translation of LIDAR data into an image input format for training a selected DL architecture. That is, all previous research used 2D images extracted from 3D LIDAR data. There has been no study of DL architectures that can work with the 3D point cloud LIDAR data directly, nor how efficacious strategies for dealing with small archaeological datasets used in the context of 2D architectures (e.g., transfer learning and data augmentation) might be when used for 3D architectures.

Our research addresses these gaps through the application of a point-based 3D shape classification DL architecture for the classification of 3D point cloud data. Specifically, the PointConv (Wu, Qi, and Fuxin 2019) DL architecture is used to classify LIDAR-based 3D point cloud data in the classification of Maya archaeological sites. The model is used in conjunction with transfer learning and data augmentation techniques to improve its efficacy on small datasets.

4. METHODS

4.1 Dataset Pre-Processing

For the experiments, we implemented 3D and 2D DL models and compared their performance. For the 3D model training, we used raw laser (LAS) formatted files annotated by the archaeologists as described in section 2. We uniformly sampled 10,024 points for each input data and computed the normal vectors from the point clouds. The primary parameters for our point cloud data include XYZ coordinates and normal vectors.

The normal vectors were calculated using a script in the open-source software CloudCompare.

In order to prepare the dataset for the 2D DL model, the hillshade image was used as shown in Figure 1. The high-resolution hillshade image was divided into smaller-size sub-images and were labeled according to the shapefile of archaeological structures described in section 2.3. Each raw image was divided into sub-images of 299×299 pixels. This size was chosen because it was large enough to include archaeological structures with background. To obtain the negative class, which does not include any archaeological structures, sub-images of the same size were used and they included a variety of natural landscape features, such as hills, mountains, and flat areas.

4.2 Data Augmentation

Large amounts of data are needed to train the deep learning models, but our dataset was not large enough, therefore, different data augmentation methods were used to create a larger and more variable dataset. For fair comparison between the 3D and 2D DL model performances, the same data augmentation strategies were employed for both models during training by randomly rotating the data and jittering both points and image pixels by adding Gaussian noise with a zero mean and 0.02 standard deviation.

The 3D model training dataset was composed of 142 positive class LIDAR data samples containing archaeological features, and 142 negative class data samples that contained natural landscape features, such as vegetation, mountains, and hills. The 2D training dataset consisted of 410 positive data samples and 430 negative class data samples with a variety of hills, mountains, and flat areas (vegetation was removed). Through data augmentation, the dataset size was tripled for both 3D and 2D model training.

4.3 Training the 3D Deep Learning Model

For our 3D implementation, the PointConv (Wu, Qi, and Fuxin 2019) model was used, which has demonstrated state-of-the-art performance in the classification of 3D point clouds. PointConv is a type of encoder-decoder framework for feature extraction and propagation that can efficiently perform convolution operations on non-uniformly sampled 3D point cloud data. The core feature extracting model takes a point cloud $P = \{p_1,p_2,...,p_L\} \in R_{L \times C_0}$ as its input, which contains L points with C_0 channels, and produces a L×C feature map at the end of its last decoder layer. PointConv utilizes a multi-layer perceptron (MLP) to learn the convolution weights on each point implicitly as a nonlinear transformation from the point coordinates to get a continuous convolution operator on irregular point clouds. Figure 3 shows how PointConv processes an example 3D point cloud input of a Maya archaeological site. In the PointConv architecture, C_{in} with C_{out} represents the dimensions of the input and output features from the K nearest neighbors. The input consists of the relative coordinates of the 3D point (P_{local}), density and characteristics (F_{in}) and relative coordinates of 3D points. After the continuous function MLP1, the corresponding feature weight W of each point is obtained, and after MLP2, the inverse density coefficient S is obtained. Finally, the input features F and output features F_{out} are calculated.

For the 3D model training, 80% of the dataset were used in training and the remaining 20% were used for testing. The

training learning rate was set to 0.001 with 32 inputs at each training batch. Stochastic gradient descent was used as an optimizer and the model was trained for 400 steps.
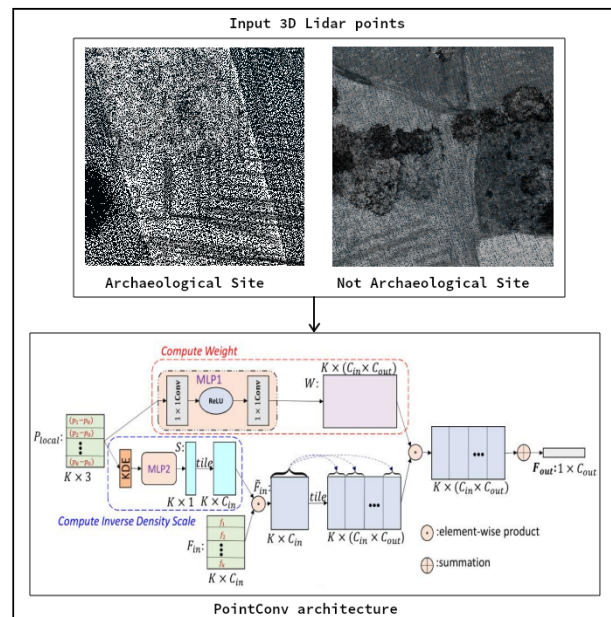

Figure 3. 3D PointConv model with input LIDAR data examples for Maya archaeological site classification is shown.

4.4 Training the 2D Deep Learning Model

The trained 2D CNN model used for the classification of archaeological sites is based on the Inception-v3 (Szegedy et al. (2016) pre-trained deep learning model. Inception-v3 can learn the most important features in different kinds of images because it is pre-trained on a large-scale hierarchical dataset called ImageNet. Given that we had a small dataset, the transfer learning approach was used to train our model. Figure 4 shows the Inception-v3 architecture along with archaeological site images.
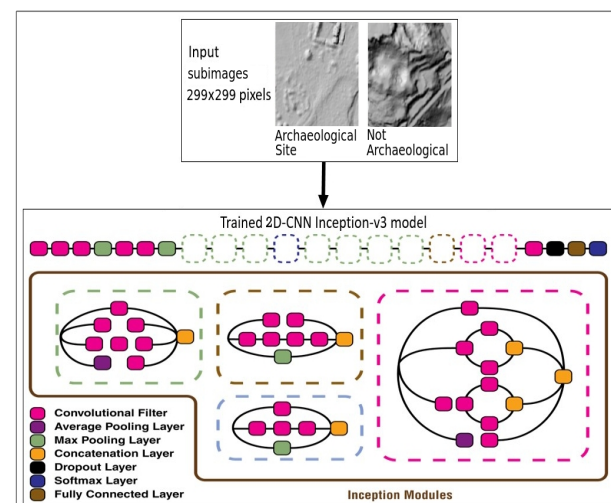

Figure 4. 2D-CNN Inception-v3 architecture is shown along with the 2D hillshade images used to train the model.

Similar to the 3D model, 80% of the annotated sub-images were used as training data for the 2D model training and the remaining 20% were used for testing. The training learning rate was set to 0.01. The training batch size was 100 and the model

was trained for 2000 epochs. The final layer of the model was retrained on our 2D high-resolution hillshade dataset and labeled using the tfClassifier repository of the Inception-v3 model (Sourcedexter, 2018).

5. RESULTS

The 3D and 2D models were evaluated based on the classification accuracy on the test datasets – which were not used in the training process. Additionally, we evaluated the models based on augmentation methods. Figure 5 shows the classification accuracy of different augmentation methods. The 3D model achieved 88% accuracy on the testing data without augmentation; 91.7% using a Gaussian noise-based approach; 92.4% using random rotations; and 95% accuracy when two different augmentation strategies were combined together – specifically, random rotation and Gaussian noise. In comparison, the 2D model was only able to achieve an accuracy of 87.8% using this same combined augmentation strategy.
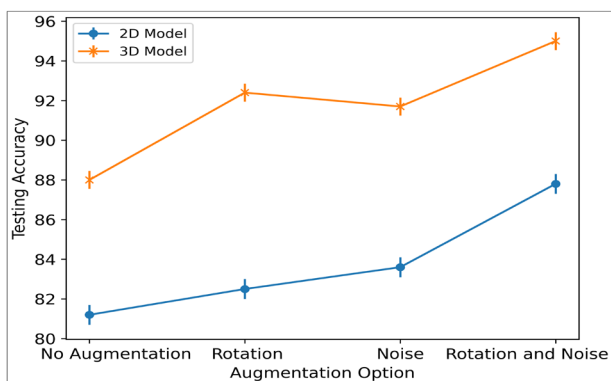
Figure 5. Classification accuracy of different data augmentation strategies for the 3D and 2D deep learning models.

The trained 3D and 2D model performances were also evaluated based on the calculation of precision (1) and recall (2).

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \qquad (1)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \qquad (2)$$

Precision measures the number of positive class predictions that actually belong to the positive class. Recall measures the number of positive class predictions made out of all positive examples found in the dataset. Using the precision and recall values, the F1 score is calculated and given in equation (3). The F1 score is the weighted average of precision and recall.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \qquad (3)$$

The 3D inference model achieved 90.5% precision and 94.1% recall values on the testing data. Based on those values, the weighted F1 value was calculated to be 92.2%. In comparison, the 2D inference model achieved 84.6% for precision and 88.9% for recall. The F1 score was 86.91%. Figure 6 shows two examples of "misclassifications" by the 3D PointConv model. For the left image, the false negative probably occurs due to landscape changes through time arising from crop cultivation. Given that the 3D training set was annotated using a shapefile of archaeological structures derived from early 1980s

pedestrian surveys, there is a high likelihood that the mounds no longer exist, and thus do not occur in the LIDAR data. Therefore, in this case the data would not actually be misclassified. As for the right image, the cause(s) of the false negative are currently unknown.
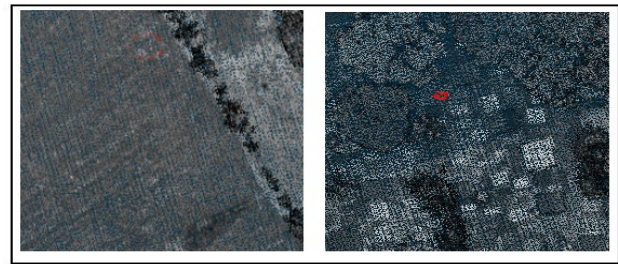
Figure 6. Examples of false negatives produced by the 3D PointConv model. The red shapes show archaeological sites that the model was unable to detect.

Figure 7 shows two examples of misclassifications (i.e., false negatives) made by the 2D CNN model. These examples illustrate the limitations of the hillshade data representation, which failed to capture archaeological ruins that were present in both of these inputs due to low mound height and/or a low ground-return density or possibly missing mounds—a subject of future research and possibly a benefit to using 3D point clouds.
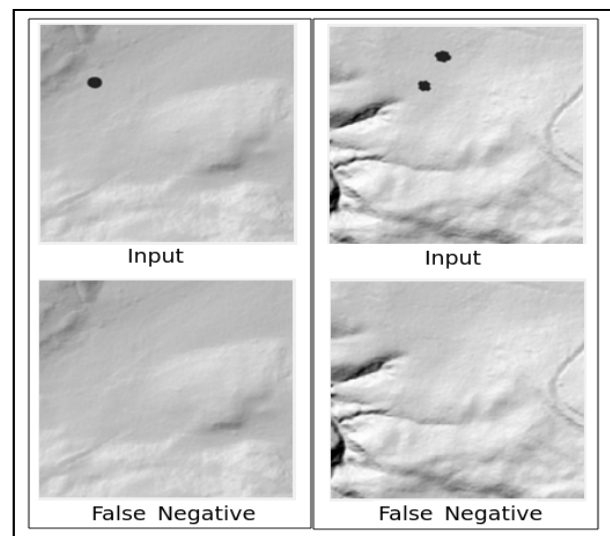
Figure 7. Examples of false negatives produced by the 2D-CNN model. The black dots show the archaeological sites that were not detected.

6. CONCLUSIONS & FUTURE WORK

While LIDAR is revolutionizing archaeological reconnaissance, archaeologists still spend an inordinate amount of time processing LIDAR data to identify archaeological sites because standard filtering algorithms often fail to capture many features (Magoni et al. 2016; Opitz and Cowley 2013). In the Maya region, identifying archaeological sites is particularly challenging because many are hidden below jungle canopy and appear as mounds that are difficult to distinguish from natural topography. To date, only a few deep learning projects have been applied to archaeology, and these have employed 2D approaches, and while several have achieved high accuracy

rates, data preparation is more time-consuming that directly using raw 3D data.

It is under these circumstances that we designed and performed this study to compare 3D and 2D deep learning approaches to classify Maya archaeological sites. Overall, the 3D model achieved the highest accuracy. This result is likely because the model is trained on 3D points, which include the z elevation (unlike the 2D hillshade images). Moreover, the data augmentation experiments showed the best accuracy when augmentation methods were combined – specifically Gaussian noise and random rotation augmentations. This combined augmentation increases the training dataset and the robustness of the model. Given the small LIDAR dataset for Copan, merging different augmentation methods using 3D data has significant potential for improving deep learning models.

While these initial results are promising, they do have some limitations. First, the DL results are dependent on the precision of the LIDAR data—areas with low point-return densities have a higher likelihood of false negatives. Second, because Maya sites exhibit diversity in architecture, infrastructure, and environment, using a small dataset does not capture all variation. For example, archaeological site locations range from low-lying scrub brush in the Yucatan Peninsula of Mexico to dense tropical forest sites in the Petén (Maya Heartland) and mountainous Highlands of Guatemala. Moreover, architectural size and style differ based on function, construction materials, and time period. Such diversity necessitates an initial large time investment in the annotation of features. Third, identification of low-lying archaeological (household) mounds depends not only on the precision of the original LIDAR data, but also on the ability to separate mounds from landscape features.

Future work seeks to overcome and/or reduce some of these limitations as well as refine the methods employed in this study. The next phase of research includes several components: (1) experimenting with additional augmentation techniques, beyond random rotation and Gaussian noise, such as applying multi-directional hillshade, PCA of hillshade, local relief, and Sky-View Factor to the DEM (Štular et al. 2012; Thompson 2020); (2), integrating raw and augmented 3D data with various 2D datasets (3) testing our methods on other Maya LIDAR datasets, and (4), identifying causes for inaccurate classification such as low mound height, vegetation types, and differential topography. Given the diversity in Maya archaeological landscapes, we plan to investigate the impact of varied approaches on different feature types in relation to landscape factors; it is likely that multiple DL methods may be required to identify and classify ancient Maya archaeological sites. To begin to investigate such differences, we are analysing the DL results in relation to other geospatial datasets such as landform and vegetation using Geographic Information Systems (GIS). Following this analysis, we will test our methods on a larger LIDAR dataset from Belize to evaluate their applicability in other Maya regions and refine them for wider-spread use for Maya archaeology.

ACKNOWLEDGEMENTS

REFERENCES

Albrecht, Conrad M, Chris Fisher, Marcus Freitag, Hendrik F Hamann, Sharathchandra Pankanti, Florencia Pezzutti, and Francesca Rossi. 2019. Learning and Recognizing Archaeological Features from LiDAR Data. 2019 IEEE International Conference on Big Data (Big Data).

Bundzel, Marek, Miroslav Jaščur, Milan Kováč, Tibor Lieskovský, Peter Sinčák, and Tomáš Tkáčik. 2020. Semantic segmentation of airborne lidar data in maya archaeology. *Remote Sensing* 12 (22):3685.

Canuto, M., Estrada-Belli, F., Garrison, T., Houston, S. et al. 2018. Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala. *Science* 28 Vol. 361, Issue 6409; DOI: 10.1126/science.aau0137

Chase, A. and Chase, D. 2017. Detection of Maya ruins by LiDAR: Applications, case study, and issues. In *Sensing the Past*, pp. 455-468. Springer, Cham.

Fash, W. 2001. *Scribes, Warriors, and Kings: The City of Copan and the Ancient Maya.* Thames and Hudson.

Garrison,T.,Houston, S., Firpi. O. 2018. Recentering the rural: Lidar and articulated landscapes among the Maya. Journal of Anthropological Archaeology 53:133-146. doi.org/10.1016/j.jaa.2018.11.005.

Girardeau-Montaut, D., 2016. CloudCompare. Retrieved from CloudCompare: https://www. danielgm.net/cc.

Guo, Yulan, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. 2020. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence.*

Guyot, Alexandre, Marc Lennon, Thierry Lorho, and Laurence Hubert-Moy. 2021. Combined Detection and Segmentation of Archaeological Structures from LiDAR Data Using a Deep Learning Approach. *Journal of Computer Applications in Archaeology* 4 (1):1.

Inomata, T., Triadan, D., Vázquez López, V.A. *et al.* 2020. Monumental architecture at Aguada Fénix and the rise of Maya civilization. *Nature* **582,** 530–533. doi.org/10.1038/s41586-020-2343-4

Kazimi, Bashir, Frank Thiemann, Katharina Malek, Monika Sester, and Kourosh Khoshelham. 2018. Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data. Proceedings of the 2nd Workshop On Computing Techniques for Spatio-Temporal Data in Archaeology and Cultural Heritage, Melbourne, Australia.

Khurana, Palak, Anshika Sharma, Shailendra Narayan Singh, and Pradeep Kumar Singh. 2016. A survey on object recognition and segmentation techniques. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. 2012. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems.

LeCun, Yann, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence Jackel, D. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1 (4):541-551.

Li, Xiang, Ling Peng, Yuan Hu, Jing Shao, and Tianhe Chi. 2016. Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research* 23 (22):22408-22417.

Magoni, Aline, Travis W. Stanton, Nicolas Barth, Juan Carlos Fernandez-Diaz, José Francisco Osorio León, Francisco Pérez Ruíz and Jessica A. Wheeler. 2016. Detection Thresholds of Archaeological Features in Airborne Lidar Data from Central Yucatán. *Advances in Archaeological Practic*e 4 (3): 232-248.

Opitz, Rachel and David Cowley (eds). 2013. *Interpreting Archaeological Topography: 3D Data, Visualisation and Observatio*n. Oxbow Books.

Piaggesi, Simone, Laetitia Gauvin, Michele Tizzoni, Ciro Cattuto, Natalia Adler, Stefaan Verhulst, Andrew Young, Rhiannan Price, Leo Ferres, and Andre Panisson. 2019. Predicting City Poverty Using Satellite Imagery. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Pires de Lima, Rafael, and Kurt Marfurt. 2020. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing* 12 (1):86.

Prufer, K. and Thompson, A. 2016. Lidar-based analyses of anthropogenic landscape alterations as a component of the built environment. *Advances in Archaeological Practice* 4(3): 393-409. Cambridge University Press.

Qi, Charles R, Li Yi, Hao Su, and Leonidas J Guibas. 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*.

Richards-Rissetto, H. 2010. Exploring Social Interaction at the Ancient Maya City of Copan, Honduras: A Multi-Scalar Geographic Information Systems (GIS) Analysis of Access and Visibility. PhD. Dissertation, University of New Mexico: Albuquerque.

Somrak, Maja, Sašo Džeroski, and Žiga Kokalj. 2020. Learning to classify structures in ALS-derived visualizations of ancient Maya Settlements with CNN. *Remote Sensing* 12 (14):2215.

Sourcedexter, 2018. tfClassifier: Tensorflow based training and classification scripts for text, images. Online: https://github.com/sourcedexter/tfClassifier

Štular, Benjamin, Žiga Kokalj, Krištof Oštir, Laure Nuninger, 2012. Visualization of lidar-derived relief models for detection of archaeological features, Journal of Archaeological Science 39 (11): 3354-3360,

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826

Thompson, Amy. 2020. Detecting Classic Maya Settlements with Lidar-Derived Relief Visualizations. *Remote Sensing*. 2020; 12(17):2838.

Verschoof-van der Vaart, W. and Lambers, K. 2019. Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology* 2 (1):31-40

Verschoof-van der Vaart, Wouter B, Karsten Lambers, Wojtek Kowalczyk, and Quentin PJ Bourgeois. 2020. Combining deep learning and location-based ranking for large-scale archaeological prospection of LiDAR data from the Netherlands. *ISPRS International Journal of Geo-Information* 9 (5):293.

von Schwerin J., Richards-Rissetto H, Remondino F., Grazia Spera, M., Auer M., Billen, N., Loos, L., Stelson, L., Reindel, M. 2016. Airborne LiDAR Acquisition, Post-Processing and Accuracy-Checking for a 3D WebGIS of Copan, Honduras, *Journal of Archaeological Research: Reports* (5):85-104. DOI:10.1016/j.jasrep.2015.11.005

Wu, Wenxuan, Zhongang Qi, and Li Fuxin. 2019. Pointconv: Deep convolutional networks on 3d point clouds. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks" Advances in neural information processing systems.

Zhao, Bo, Jiashi Feng, Xiao Wu, and Shuicheng Yan. 2017. "A survey on deep learning-based fine-grained object classification and semantic segmentation." *International Journal of Automation and Computing* 14 (2):119-135.