Extracting Ancient Maya Structures from Aerial LiDAR Data using Deep Learning

Fatema-E- Jannat, Jincheng Zhang, Andrew Willis
Dept. of Electrical and Computer Engineering
University of North Carolina at Charlotte
Charlotte, North Carolina, USA
{fjannat, jzhang72, arwillis}@uncc.edu

William Ringle
Dept. of Anthropology
Davidson College
Davidson, North Carolina, USA
biringle@davidson.edu

***Abstract*—The advent of LiDAR technology has had a revolutionary impact on archaeological prospection by vastly enlarging the coverage of ancient landscapes and consequently the number of ancient surface features. However, manual analysis by experts requires a significant time and money investment. This paper describes a deep learning model developed to segment, i.e., label, the semantics of objects of interest as a means to augment or supplant manual labeling of LiDAR data. The U-Net deep learning model forms the backbone of the system which has shown success in providing accurate outputs on similar LiDAR data set. The trained U-Net model is integrated into an inference pipeline to transform expansive LiDAR datasets into labeled output images. Work focuses on the classification of two semantic types: (1) platforms and (2) annular structures whose attributes, e.g., location, shape, and distribution, play an important role in improving our understanding of ancient Maya civilizations. This article provides a deep learning-based system that efficiently extracted these structures. CNN-generated inferences were compared against expert-labeled features to measure algorithm performance. Results for a LiDAR survey of 479 sq. km. indicate that the CNN provides an IoU performance of 0.82 and 0.74 for annular structures and platforms respectively. The discussion further analyzes how IoU performance relates to the viability of this approach as an aid or substitute for manual labeling.**

***Index Terms*—LiDAR, remote-sensing, segmentation, deep-learning, U-Net**

I. Introduction

Accurate and extensive knowledge of ancient landscapes is central to our ability to reconstruct ancient settlements and their surroundings, allowing us to make inferences concerning demography, economic activities, and sociopolitical organization. Traditional mapping methods are costly and labor intensive, however, especially in heavily vegetated areas such as the Maya lowlands. This has been revolutionized by LiDAR (Light Detection and Ranging), a remote sensing technique that creates high-resolution elevation models of the earth's surface using a laser scanner. Laser point clouds can then be converted to Digital Terrain Models (DTMs) for the purpose of identifying features of interest.

Detailed imagery of ancient settlements over very large regions can thus be had in a small fraction of the time pedestrian coverage would require, but comes with a sharp increase in the time needed to manually identify features in the imagery. As data volume increases, speeding up and automating the annotation of features has become an active area of research. In this paper, we report on the application of deep-learning segmentation to LiDAR data from the Puuc region of Yucatan, where ancient Maya communities flourished from 500 BC to AD 1000.

Deep learning-based semantic segmentation has been used for an automatic annotation system of objects of interest with the model's objective being to assign a class label to each pixel in the image. The goal is to train the model to be able to segment every pixel in any input image into corresponding class labels or backgrounds.

Platforms and annular structures built by the Maya of the Puuc region are the focus of this paper's LiDAR data analysis (see Fig. 1). (Stone platforms supported houses and administrative buildings while annular structures were open ovens probably used for lime production.) A small portion of the region was manually labeled and verified by experts. The goal is to automatically annotate the remaining portion of the region using a deep learning (DL) model trained with that small subset of labeled data.

The data set includes LiDAR data in tiff format and expert-drawn training polygons of annular structure and platforms provided as shapefiles. We converted the raw data, which was provided in tiff and shapefile format, into a format that machine learning models could understand. In addition, we take the massive LiDAR data set and divide it into small tiles of 128×128 to make a suitable input for our model. The segmentation model's output measures 128×128 as well. However, this format of output gives little information to archaeologists. To provide a more useful output format, we stitched all the output tiles together to produce the inferred output at the same size as the original input size but with a predicted masked region showing the objects of interest.

DL models have become extremely popular in recent years across a variety of computer vision fields due to their capacity to automatically extract features from images while boosting accuracy and productivity. Nevertheless, putting these DL models into practice calls for specialized programming and machine learning expertise, which are uncommon in archaeology. U-Net [1] was used to segment archaeological structures in LiDAR data with the intent to reduce manual labeling of these data.

978-1-6654-7611-9/23/$31.00

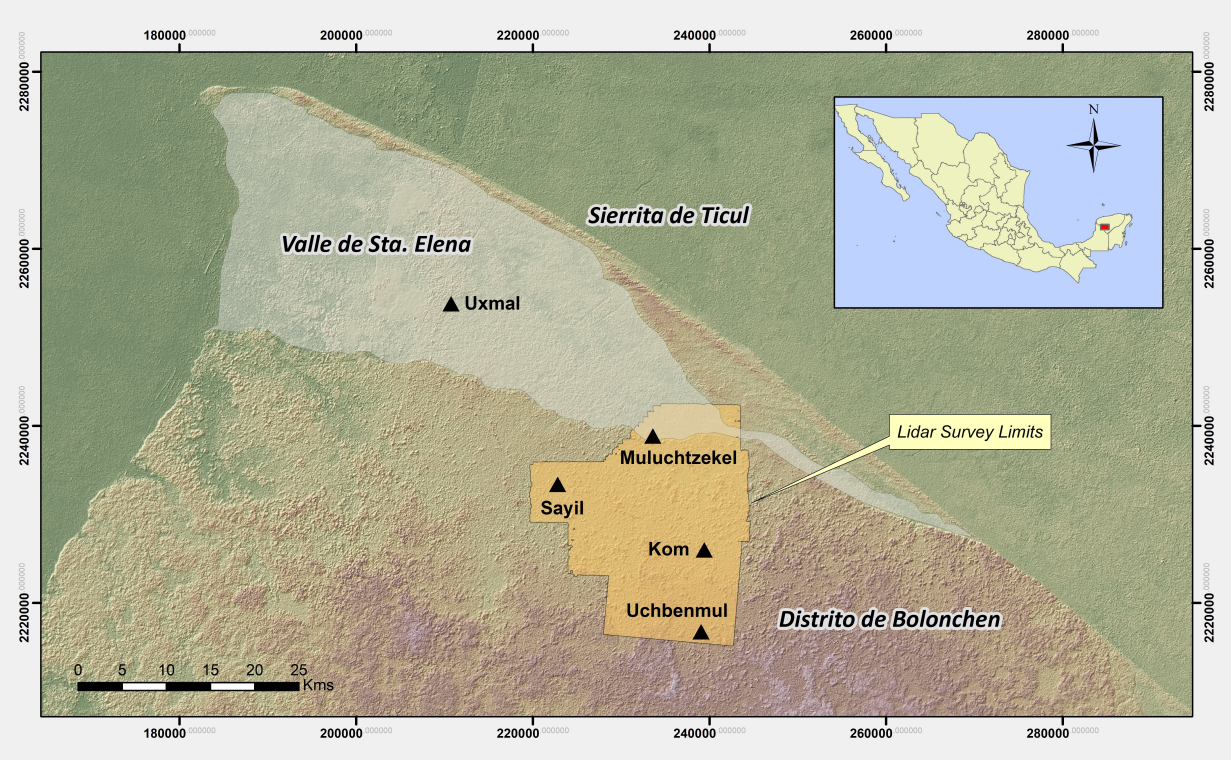

Fig. 1. Location map of the research area - the Puuc region of northern Yucatan, Mexico.

The main contributions of this paper are summarized as follows:

- the model demonstrates the performance of 0.82 IOU for annular structures, a type of structure analyzed for the first time by means of DL.
- an image augmentation pipeline is proposed that leverages a relatively small amount of labeled LiDAR data and produces results competitive with the recent works for platform structures (0.74 IOU).
- additional archaeological analysis is provided that measures how effective this tool is in practice. Findings indicate the tool performs satisfactorily for identifying annular structure, but more work is required for comparable utility regarding platform identification.

II. Related Works

In the field of computer vision, deep learning models currently dominate thanks to consistent improvements in accuracy and speed for tasks such as image classification and object detection. LeCun's creation of CNN [2] in the early 1990s opened up new possibilities for image classification. However, the revolution in CNN came in 2012 with Krizhevsky's 'AlexNet' [3], which classified 1.2 million images into 1000 classes with an error rate of 15.3%. CNN continues to advance with the development of ResNet [4], GoogleNet [5], SqueezeNet [6], DenseNet [7].

Despite the enormous success of deep learning methods in other computer vision fields, the application of deep learning in archaeology is still in its early stages. This can be attributed, in part, to the difficulty of implementing deep learning models, which calls for a high level of computer science proficiency. The limited availability of data for archaeological sites may also be another factor. However, in recent years, with the advent of LiDAR technology and steady improvement in the predictive capabilities of DL models, interest in the application of DL in archaeology has grown. A considerable amount of work has been carried out using remote sensing imagery, primarily for classification tasks [8]–[14].

In 2022 Banasiak, P.Z. [15] implemented deep learning neural networks (DLNN) for the automatic recognition of archaeological monuments in the Polish part of the Białowieża Forest. For this, they used Airborne Laser Scanning (ALS) data and performed semantic segmentation using the U-Net [1] model. They were able to achieve IoU values for ancient field system banks, ancient field system plots, and burial mounds of 0.41, 0.616, and 0.62 respectively.

Using several variations of the VGG-19 Convolutional Neural Network (CNN) applied to Airborne Laser Scanning (ALS) data from the Chact'n region, Somrak, M. [16] showed that a CNN model can classify different types of ancient Maya structures. Concurrently, Bundzel, M. [17] used Pacunam Initiative LiDAR survey data of the lowland Maya region in Guatemala to identify the locations of ancient construction activity and the remains of ancient Maya buildings. They performed semantic segmentation using two different deep learning models, U-

Net [1] and Mask R-CNN [18] and demonstrated that they were able to identify 60–66% of all objects and 74–81% of medium-sized objects using U-Net.

Using LiDAR data from the Maya forest region, Landauer, J. [19] trained an ensemble of DeepLabV3+ [20] and HRNet [21] network to automatically detect reservoirs, buildings, and platforms, achieving an average IoU of 0.8275 across all three classes.

It is evident from all the published articles that DL is becoming more and more popular in archaeology. Its application in classifying LiDAR imagery and detecting archaeological features reduces the need for manual inspection while saving time and money.

III. Methodology

Our method to develop the proposed system started with unlabeled LiDAR data collected by the National Center for Airborne LiDAR and Mapping (NCALM) from the Puuc region of Mexico (see fig. 1 at a resolution of 0.5 m./pixel. Platform and annular structures were then labeled manually to generate a ground truth label set. Due to the large size of the raw LiDAR data set, it was decomposed into tiles to accommodate training a U-Net model. Due to the small number of available labels a significant amount of data augmentation was used to boost the performance of the trained U-Net model. This section also describes other numerical considerations that served to boost performance which include: (1) normalization of the elevation data and (2) selection of an appropriate loss function. Inference using the CNN was also customized by merging multiple classifications for each pixel into a final classification where each classification considered different regions in the vicinity of the pixel.

A. Ground Truth Labeling

The procedure for generating ground truth data follows the standard practice for developing deep learning systems. Specifically, ground truth labels were specified manually as polygons that enclose each object of interest as shown in Fig. 3(b). Binary masks were derived from polygon data to generate ground truth labels for tiles during training. Fig. 3(a) shows a sample visualization of the elevation data using a hillshade algorithm to highlight variations in elevation. Fig. 3(b) shows ground truth polygon region labels for platforms (green) and annular structures (yellow) superimposed labeling over the hillshade version of the data. Fig. 3(c) and (d) show examples of binary mask representations of the annular structures and platforms, respectively.

B. Model Architecture

U-Net is an encoder-decoder type neural network architecture proposed by O. Ronneberger et al. [1]. This work adopts the U-Net neural architecture to segment the LiDAR data. Although this U-Net was primarily developed for the purpose of segmenting biomedical images, it has demonstrated excellent success in segmenting remotely sensed images including LiDAR images [15], [17], [22].

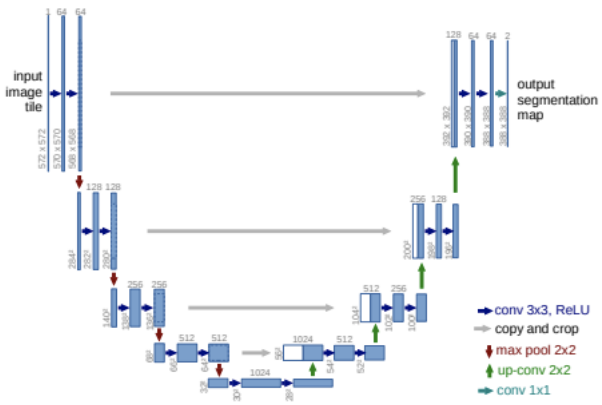

Fig. 2. U-Net architecture [1]

U-Net consists of two parts, an encoder, and a decoder. The encoder is used to down-sample the input, and the decoder is used to up-sample it. Skip connections concatenate encoder feature maps to the decoder feature maps in between encoders and decoders, allowing information to be passed from an earlier stage to a later stage directly and significantly reducing the issue of vanishing gradient. An illustration of U-Net architecture from the original paper [1] is shown in Fig. 2.

C. Augmentation

Due to the limited number of ground truth labels of our classes, data augmentation methods were applied to generate samples of alternate realizations of our classes in image tile data. Data augmentation was performed via two image sampling approaches: (1) random background sampling and (2) random rotations and translations of each training label within the tile perceptual field. The first sampling approach focuses on the development of a comprehensive model of background by collecting random samples that do not relate to the location of labeled structures. The second sampling approach ensures that the small number of existing labels is also sampled with sufficient variation and density to maximize the performance of the trained network. In both cases, regions of the LiDAR source data equivalent to the input layer size of our network were extracted with varying positions and orientations.

D. Numerical Considerations

1) Normalization: We normalize the range of elevation data within each training tile to the $[0, 1]$ interval which has proven to improve classification results as noted by other researchers [22]. The normalization technique used is shown in equation (1)

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X denotes an elevation value, X_{max}, X_{min} denote the maximum and minimum elevation values in the tile, and X_{norm} denotes the normalized elevation result.

2) Loss Function: The standard loss function for binary classification is used which is the categorical cross-entropy loss function as shown in equation (2)

$$H(p,q) = H(p) + D_{\mathrm{KL}}(p \parallel q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x) \quad (2)$$

where $H(p)$ denotes the entropy of the random variable p, $D_{\mathrm{KL}}(p \parallel q)$ denotes the Kullback-Leibler divergence of p from q and x is an event drawn from the probability space of these random variables $\mathcal{X}$. The Keras cross entropy also provides a smoothing parameter which serves to penalize overconfident outputs and helps prevent over-fitting [23]. Results shown in this article use a smoothing parameter of 0.1.

E. Inference of Structure Labels

Our method for inference differs from the standard approach for CNN inference. Specifically, our approach decomposes the unlabeled data into 128×128 input tiles where these tiles may overlap. This modification allows the inferred output classifications to take into consideration image values from a larger perceptual field and also reduces adverse impacts associated with labels that lie on the boundary of the image. We refer to this approach as a sliding-window classification output, and our results consider sliding window skips of 32 pixels horizontally and vertically. For our chosen tile size, this results in classifying each pixel 16 times. The results are combined by choosing the class having the highest joint probability across all tile classifications.

IV. DATA SET

A. Data Acquisition

LiDAR imagery was provided by the Bolonchen Regional Archaeological Project (BRAP), directed by Ringle, Tomás Gallareta Negrón (INAH/CY), and George Bey (Millsaps College), and was collected by NCALM with funding from the National Science Foundation. Most of the coverage was obtained in 2017 and then supplemented by areas to the east and west in 2022. Our coverage also borders a data set to the south collected for forestry research by the Alianza MexicoREDD+, which has courteously allowed the use of their information. In all, our data set comprises $478.68km^2$ of continuous coverage. (Note that this is an increase from that reported in [24]). LiDAR data was obtained using a Teledyne Optech TitanMW(14SEN/CON340) sensor mounted in a small airplane flying at an altitude of 600-650 m. About 62.6% of our pulses produced ground returns, resulting in a density of 10.6/m2. From this, a 0.5 m DTM raster was produced that forms the input for our analysis. (The Alianza point cloud, originally used to produce a 1.0 m-density raster, was reclassified and resampled to produce a DTM at the same resolution).

B. Data Regions

We selected data from three sites (Muluchtzekel, Kom, and Uchbenmul) and a region of $67.6km^2$ centered around the site of Sayil (Fig. 1). The first three had been partially covered by the ground survey and so were useful for training purposes. The Sayil region has not been explored by the BRAP project, although much of the site of Sayil was mapped in the 1980s.

1) Muluchtzekel: Muluchtzekel(MLS) is the largest site we have ground-surveyed and covers about $3.8km^2$. It lies at the interface between the relatively flat lands to its north and the Bolonchen Hills behind it.

2) Kom: Kom is a second-rank site within the Bolonchen Hills. Although much smaller than Muluchtzekel, it does possess a palace and several elite residential platforms.

3) Uchbenmul: Uchbenmul (UCB) is the farthest south of the sites and also lies within the Bolonchen Hills. It lacks a palace (though it has a much earlier acropolis) and has only a few vaulted structures and so represents the lower tier of regional sites.

4) Sayil: The Sayil block is named for the principal site within it but includes a number of other sites.

A detailed summary of the ground survey statistics for these regions is provided in Table.I.

V. RESULTS

A. Experiments

Our experimental results consider 4 data sets from the three sites and the Sayil region mentioned above. We refer to them as: (1) Kom, (2) MLS, (3) UCB and (4) Sayil. Kom, MLS, and UCB were partially labeled and Sayil was unlabeled. As shown in Table I, the Kom and MLS data sets include labels for annular structures and platforms, while UCB includes labels only for annular structures. Two U-Net segmentation models were trained: one for platforms and the other for annular structures.

1) Label Generation and Data Augmentation: A total of 96 annular structures were manually ground-truthed and used for training (see Table I). The platform segmentation algorithm was trained with the Kom and MLS data sets, which contain 861 labeled platforms. Due to the small amount of labeled data, augmentation methods were used extensively. Augmentations consist of a random translation and rotation of original LiDAR elevation data. In our case, this consisted of 1000 random augmentations, where random tiles of the data were sourced, and 150 augmentations of regions in the vicinity of each training label. The resulting collection of augmented imagery was subdivided into training, validation, and testing data sets where the number of images in each set was 65%, 15%, and 20% respectively of overall number of images.

B. Training the U-Net Model

Experiments were conducted using an NVIDIA RTX A6000 GPU with CUDA 12.0 for training the U-Net. Training parameters included choosing a batch size of 100 images and a learning rate of 0.001 with the Adam optimizer. Training

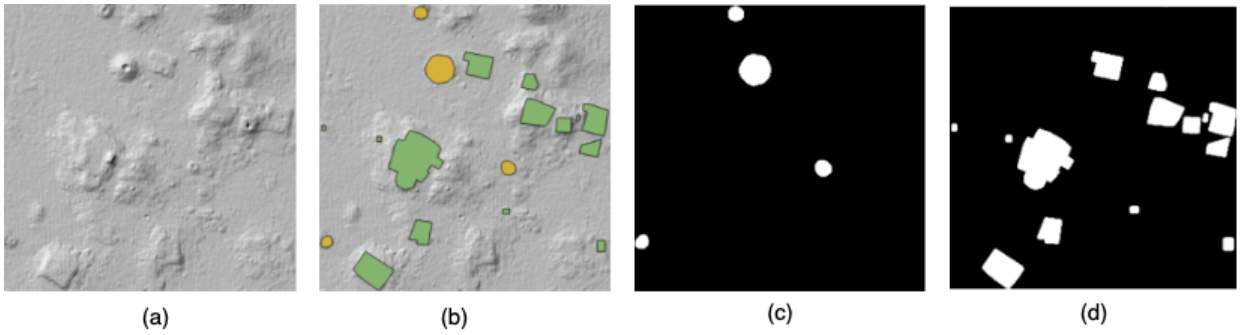

Fig. 3. Visualization of input, ground truth labeling, and corresponding masks for the object of interest. (a) shows a hillshade visualization of the measured elevation data, (b) shows ground truth polygon labels for annular structures (yellow) and platforms (green) superimposed over the hillshade data, (c) and (d) show the binary masks for annular structures and platforms in (b).

TABLE I
GROUND SURVEY STATISTICS FOR THREE SITES.

Site	Overall Area	Area Surveyed (ha.)	Platforms	Annulars
Muluchtzekel	600.00	403.0	613	54
Kom	125.30	90.7	248	32
Uchbenmul	27.70	31.1	–	10

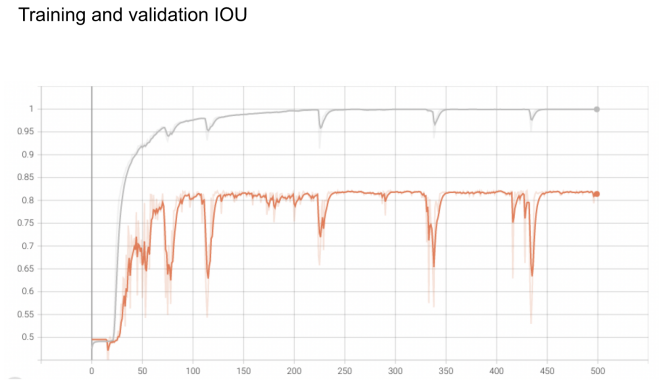

Fig. 4. IoU curve for training and validation set with U-Net model for binary segmentation of Annular Structures.

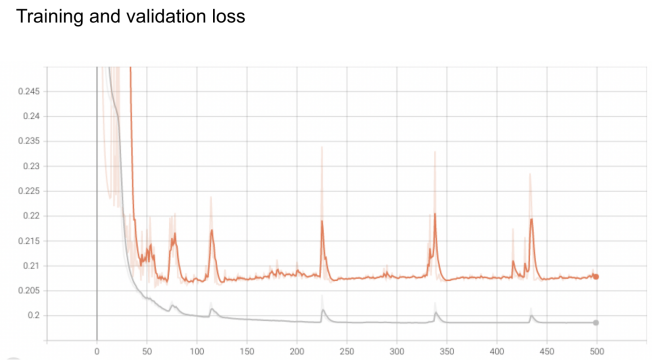

Fig. 5. Loss curve for training and validation set with U-Net model for binary segmentation of Annular Structures.

optimization was run for 500 epochs and required approximately 4 hours to complete. The training and validation IoU and loss curves for the annular structure segmentation model are depicted in Fig. 4 and Fig. 5 where the number of epochs is denoted on the x-axis and the IoU or loss is measured on the y-axis. Fig. 6 and Fig. 7 show the training and validation IoU and loss curves for the platform segmentation model.

C. Evaluation Metrics

We used the IoU (Intersection over Union) metric, which calculates the amount of overlap between two masks of ground truth and prediction, to assess the model's performance. The IoU value ranges from 0 to 1, with 1 denoting perfect overlap and 0 denoting imperfect overlap. The formula for IoU is given in equation (3).

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

Here, the number of shared pixels in both ground truth and prediction images is represented by the intersection of $(A \cap B)$, and the total number of pixels from both images is represented by the union of ground truth and prediction $(A \cup B)$.

1) Annular Structure Classification: The model converges and archives 0.82 IOU on validation data, as seen in Fig. 4. Fig. 8(b) displays an inference on a small portion of the MLS region and depicts how well it extracts annular structures. Regions marked in light purple indicate True Positive results where the label was correctly assigned. False Positives are indicated by light blue; the model labeled them incorrectly. Regions marked in light pink indicate correctly inferred classifications.

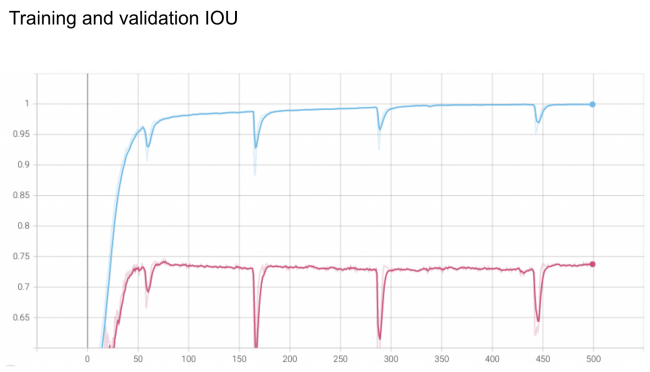

Fig. 6. IoU curve for training and validation set with U-Net model for binary segmentation of Platforms.

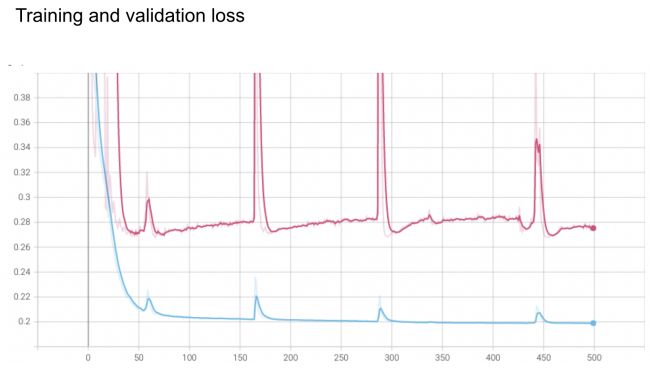

Fig. 7. Loss curve for training and validation set with U-Net model for binary segmentation of Platforms.

2) Platform Classification: The model for platform classification succeeds in achieving 0.74 IoU on the validation data (Fig. 6). To demonstrate how well platforms are extracted, Fig. 8(a) shows an inference on a small portion of the MLS region. These segmentation results are consistent with the most recent findings [19], [25] (also 0.74 IoU). Given the disparity in labeled source data between this work and [19], [25](they had approximately 951 images of platforms for their training), this is an indication that our augmentation provides sufficient variability to generate a state-of-the-art platform classifier.

D. Archaeological Analysis

Archaeological analysis reviews classifications performed on the MLS and Sayil data sets to evaluate the utility of this tool in practice. Selected classification results are shown in Fig. 8(a,b) and Fig. 9. These results were also analyzed manually to evaluate the benefit of this tool as an aid or substitute for manual labeling.

1) Annular Structures: The chosen excerpt shown in Figure 9 depicts the four possible classification outcomes: (1) True Positive, (2) True Negative, (3) False Positive, and (4) False Negatives. Fig. 9 shows that the system is capable of extracting annular structures (marked as True positive), and rejecting similar shapes which are typically chultuns (Maya cisterns) and other anomalies, e.g., pits (marked as True Negative). Fig. 9 also shows classification errors which include classifying a pit and a pyramid with a looter's pit at its peak incorrectly as annular structures (pointed as False Positive). Some annular structures were missed (marked as false negatives). It is also important to note that classification results detected 13 annular structures in the Sayil data set that were overlooked in the initial manual review. The performance of this classifier was found to be appropriate for accelerating manual labeling.

2) Platform Structures: Fig. 8(a) shows inference results for the MLS region. Correct classifications are valuable as they provide both location and shape of the structure. Shape features are particularly valuable as they can be used to infer structure type and the energy involved in their construction, but they are time-consuming to manually specify.

These results also show some inaccuracies in classification results. The frequency and regularity of these errors lead to the conclusion that despite the performance report of 0.74 IoU, this level of performance can aid manual labeling but is not sufficient to act as a substitute. Therefore, further investigation into this topic is required to create automated tools that use deep learning to consistently create pointwise features for platforms.

VI. Conclusions

In this article, we have investigated the efficacy of applying a deep learning model to extract ancient Maya structures from LiDAR data. Deep learning models are being applied increasingly in several computer vision applications but there has been very little work on extracting archaeological features from LiDAR data with deep neural networks. This paper describes a system that successfully extracts annular and platform structures using the U-Net model with an IoU performance of 0.82 for annular structures and an IoU performance of 0.74 for platform structures. After analyzing the result, our finding indicates that these are sufficiently accurate identifications when supplemented with manual revision. We were pleased to find that some structures missed during preliminary manual labeling were extracted by the model. Additionally, processing the errors took less time than it would if the deep learning model had not been used and the process had to be completed manually.

Acknowledgments

We would like to acknowledge funding for LiDAR acquisition from the National Science Foundation (Award No. 1660503) and the National Center for Airborne LiDAR Mapping for the collection of the data used in this article.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

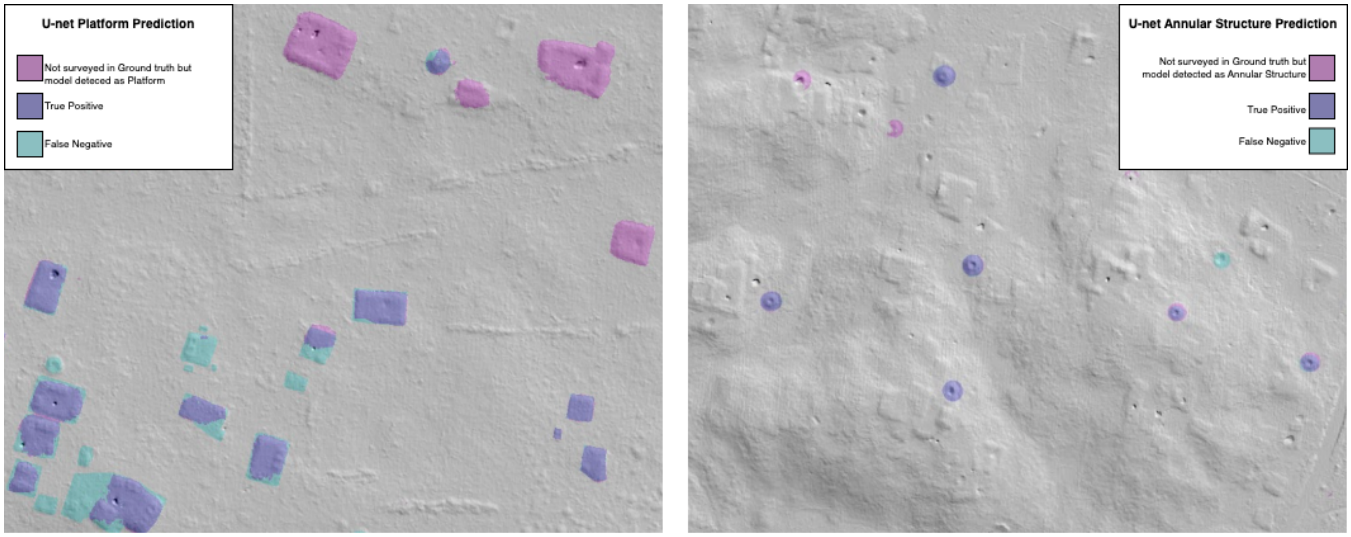

Fig. 8. Visualization of inference on a small part of MLS region from the trained model. (a) is from Platforms segmentation and (b) is from Annular Structure segmentation.

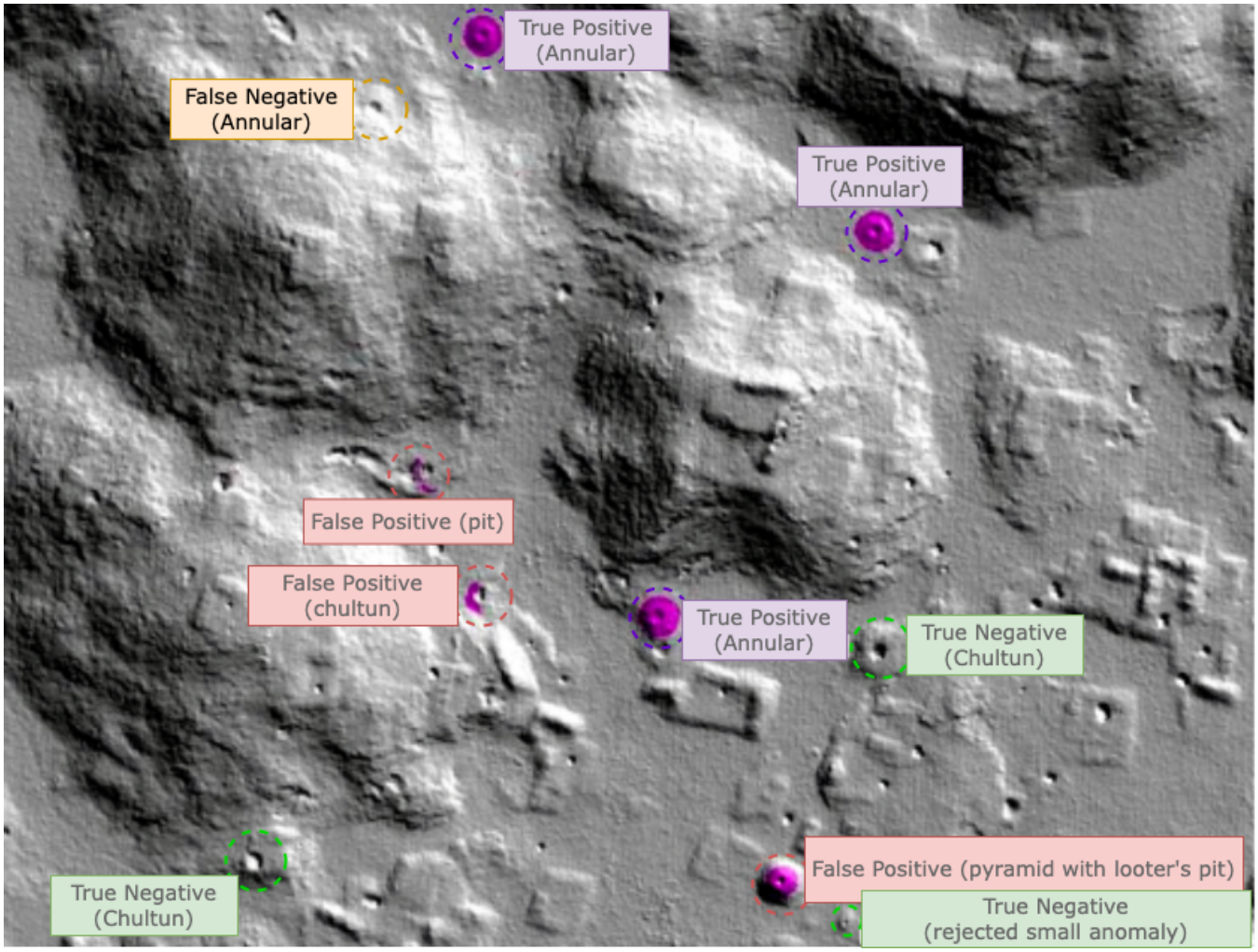

Fig. 9. Visualization and analysis of inference on a small part of Sayil region.

[2] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, contour and grouping in computer vision*. Springer, 1999, pp. 319–345.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[6] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, 2016.

[7] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jul 2017, pp. 2261–2269.

[8] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 204–207.

[9] M. Wang, X. Zhang, X. Niu, F. Wang, and X. Zhang, "Scene classification of high-resolution remotely sensed image based on resnet," *Journal of Geovisualization and Spatial Analysis*, vol. 3, no. 2, 2019.

[10] M. Schmitt and W. Y-L, "Remote sensing image classification with the sen12ms dataset," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. -2-2021, pp. 101–106, 2021.

[11] F.-E. Jannat and A. R. Willis, "Improving classification of remotely sensed images with the swin transformer," in *SoutheastCon 2022*, 2022, pp. 611–618.

[12] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[13] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sensing*, vol. 13, no. 3, 2021.

[14] Y. Bazi, M. M. Al Rahhal, H. Alhichri, and N. Alajlan, "Simple yet effective fine-tuning of deep cnns using an auxiliary classification loss for remote sensing scene classification," *Remote Sensing*, vol. 11, no. 24, 2019.

[15] P. Z. Banasiak, P. L. Berezowski, R. Zapłata, M. Mielcarek, K. Duraj, and K. Stereńczak, "Semantic segmentation (U-Net) of archaeological features in airborne laser scanning," *Remote Sensing*, vol. 14, no. 4, 2022.

[16] M. Somrak, S. Džeroski, and Kokalj, "Learning to classify structures in als-derived visualizations of ancient maya settlements with cnn," *Remote Sensing*, vol. 12, no. 14, 2020.

[17] M. Bundzel, M. Jaščur, M. Kováč, T. Lieskovský, P. Sinčák, and T. Tkáčik, "Semantic segmentation of airborne lidar data in maya archaeology," *Remote Sensing*, vol. 12, no. 22, 2020.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[19] J. Landauer, B. Hoppenstedt, and J. Allgaier, "Image segmentation to locate ancient maya architectures using deep learning," *Publishers Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia*, p. 7, 2022.

[20] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 833–851.

[21] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5686–5696.

[22] M. Hliboký, M. Bundzel, and S. Husár, "Detecting ancient mayan construction activity by U-Net," in *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2022, pp. 231–236.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[24] W. M. Ringle, T. Gallareta Negrón, R. May Ciau, K. E. Seligson, J. C. Fernandez-Diaz, and D. Ortegón Zapata, "Lidar survey of ancient maya settlement in the puuc region of yucatan, mexico," *PLoS One*, vol. 16, no. 4, pp. 1–54, 2021.

[25] T. Hellweg, S. Oehmcke, A. Kariryaa, F. Gieseke, and C. Igel, "Ensemble learning for semantic segmentation of ancient maya architectures," *Publishers Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia*, p. 13, 2022.