To the University of Wyoming:

The members of the Committee approve the thesis of Ethan Dowalter presented on April, 30 2024.

Lars Kotthoff, Chairperson

Shannon Albeke, Outside Member

Mike Borowczak

Jeff Hamerlinck

,

APPROVED:

Bryan Shader, Head, Electrical Engineering and Computer Science

Cameron H. G. Wright, Dean, College of Engineering and Physical Sciences

Dowalter, Ethan, *Binary Classification on Columnar Cacti in Point Clouds Using Neural Networks*, M.S., Electrical Engineering and Computer Science, May, 2024.

This study investigates the application of a neural network model for the binary classification of point cloud data into cactus and non-cactus categories, aimed at facilitating ecohydrological research by simplifying cactus volume and surface area calculations. The most effective model found used a multilayer perceptron (MLP) architecture, though using a one-dimensional convolutional neural network (1D CNN) architecture proved only marginally less effective. Utilizing a 10-fold cross-validation method, the models demonstrated variable accuracy and recall across datasets, highlighting the challenges of classifying spatial data, though both models showed promising results. This approach offers a foundational step towards leveraging neural networks in environmental studies, demonstrating promise in the possibility of greatly accelerating the process of measuring columnar cacti on a landscape-scale to monitor the health of aridland ecosystems.

BINARY CLASSIFICATION ON COLUMNAR CACTI IN POINT CLOUDS USING NEURAL NETWORKS

by

Ethan Dowalter, B.S. Computer Science

A thesis submitted to the
Electrical Engineering and Computer Science
and the
University of Wyoming
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE
in
COMPUTER SCIENCE

Laramie, Wyoming
May 2024

Dedicated to Holli Prichard...

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my thesis committee members, Shannon Albeke, Lars Kotthoff, Mike Borowczak, and Jeff Hamerlinck for their guidance, support, and expertise throughout the entirety of this research journey. Shannon's mentorship and expertise were instrumental in shaping this thesis. I am thankful for Lars's machine learning insights and Mike originally serving as my advisor. Though Mike decided to move onto the University of Central Florida, I am so grateful for the year we did have and the funding which has kept me afloat the past two years. Additionally, Jeff acted as yet another pivotal voice for enriching the scholarly discourse.

Furthermore, I extend my appreciation to my family and friends for their unwavering support, understanding, and encouragement during this challenging journey. Their patience, love, and encouragement have sustained me through the ups and downs of graduate school. Thank you to everyone who has played a part in this journey.

Ethan Dowalter

University of Wyoming

May 2024

Chapter 1

Introduction

The utilization of point cloud data has emerged as a transformative approach in various fields, including architecture, archaeology, construction, land surveying, and natural resource management. Point cloud data provides a comprehensive three-dimensional representation of objects and terrain surfaces, facilitating detailed analysis and decision-making. Within this realm, the detection and classification of specific objects, such as columnar cacti, present unique challenges and opportunities. Ecologists studying aridland ecosystems can use the physical dimensions of columnar cacti to effectively monitor the health of the ecosystem, being that these cacti have a disproportionately large effect on the ecosystem relative to their abundance. With the volume of a cactus, the amount of water a cactus stores can be deduced. Additionally, with the surface area of a cactus, the photosynthetic potential as well as the transpiration rate can be deduced. Traditional methods of gathering these values using hand-measurements are slow and labor-intensive. If analysis is to be done on a landscape-scale, point clouds and their subsequent analysis present a promising new way to gather the volumes and surface areas of columnar cacti.

1.1 Prior Work

1.1.1 Comparing Photogrammetric and Classic Means for Volume Estimation

This research builds off of prior work in which S.E. Albeke, D. G. Williams, and K. R. Hultine sought to compare the efficiency of estimating columnar cactus biomass via standard allometric scaling, in which the cacti are physically measured, and via photogrammetric point clouds, where the point clouds are gathered and subsequently processed inside of compatible software. In this case, Agisoft Metashape Professional was used to both build the point clouds and manually isolate the cactus bodies to then calculate their volumes and surface areas. That work demonstrated the validity of remotely captured estimates, as they matched closely with the hand-measurements. Additionally, the image capture methods were found to be approximately three times more rapid than hand measurements for the same plants, and required fewer people to collect the data. However, there was an additional overhead cost of 10 to 30 minutes for each plant for the necessary image processing involved in building the point clouds.

Plants	Difference Height m	Difference Area m^2	Difference Volume m^3	Stems
CG_01	0.140	0.207	0.019	1
CG_02*	0.116	0.054	0.003	1
CG_03	-0.260	-0.476	0.080	1
CG_04*	0.140	0.041	0.001	1
CG_05	0.128	0.100	0.003	1
OP_01	0.110	0.298	0.004	3
OP_02*	0.085	-0.137	0.003	1
OP_03	0.350	0.665	0.011	2
OP_04	0.080	0.983	0.008	9
OP_05	0.070	0.771	-0.002	6

Table 1.1: Measured difference between Hand and Photo (hand – photo) methods of five plants for each species. CG represents saguaro (Carnegiea gigantea) and OP represents organ pipe (Stenocereus thurberi). * indicates plants measured with Pixel 6 mobile phone with the remaining plants using DJI Mavic Pro.

1.1.2 Workflow for Volume Calculation

Once the photos are gathered and used to generate the point clouds, traditionally, the next step would require manually selecting points that correspond solely to the cacti, separating them from other objects and noise within the dataset. After the cacti points are segregated, the next step involves creating a 3D mesh. This mesh is a polygonal model constructed from the points, effectively forming a digital skin over the point cloud. Following the initial construction of the mesh, the holes at the bottom of the cacti, where the stem meets the ground, must be filled in to form a closed shape. Finally, with any suitable point cloud software, like Agisoft Metashape Professional, one can use a built-in function to calculate the volume and surface area of a closed mesh with a single click.

1.2 Background

This paper sits at the intersection of four concepts which have never been combined as they are in this work: binary classification, columnar cacti, point clouds, and neural networks. As such, it is important that the reader understands each of these concepts individually before understanding how it is that they are combined in this work.

1.2.1 Binary Classification

Binary classification is a task that involves categorizing data into one of two distinct classes. This approach is used to answer yes/no questions, where each input (or instance) is classified into one of two predefined categories. Common examples of binary classification include detecting spam emails (spam or not spam), diagnosing diseases (sick or healthy), and predicting binary outcomes such as pass/fail or win/lose. The process involves training a model on a dataset where the outcomes are already known; this dataset is called the training set. The model learns to make predictions based on the features of the data, aiming to generalize from the training data to accurately predict the class of new, unseen instances. This is achieved through various algorithms such as logistic regression, support vector machines, decision trees, and neural networks, each providing different mechanisms for handling and classifying input data.

In this context, the two categories are often referred to as the "positive" and "negative" class instances. The positive class is typically the condition or outcome the model is particularly looking for, and the negative class represents the absence of the condition. In the case of this study, the positive class is represented by the "cactus" classification, and the negative class is represented by the "non-cactus" classification. To evaluate the performance of a binary classifier, a binary confusion matrix is often used (see Figure 1.1). It includes four outcomes: true positives (TP), where the model correctly predicts the positive class; true negatives (TN), where the model correctly predicts the negative class; false positives (FP), where the model incorrectly predicts the positive class; and false negatives (FN), where the model fails to predict the positive class.

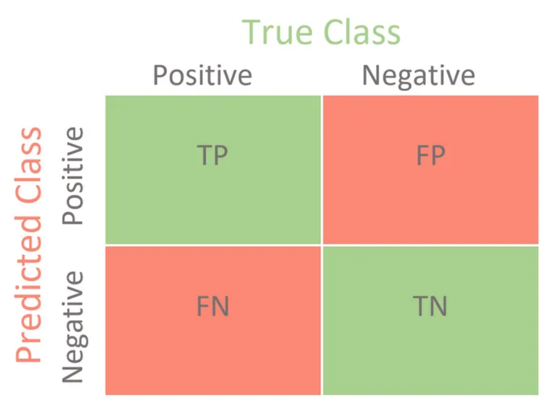

Figure 1.1: A 2x2 binary confusion matrix showing the four possible outcomes for when a model is performing binary classification. TP stands for "true positive" and indicates that a positive class instance was correctly classified. FP stands for "false positive" and indicates that a negative class instance was incorrectly classified as being positive. FN stands for "false negative" and indicates that a positive class instance was incorrectly classified as being negative. TN stands for "true negative" and indicates that a negative class instance was correctly classified.

1.2.2 Columnar Cacti

Columnar cacti, characterized by their cylindrical stems and spiny structures (as seen in Figure 1.2), play a vital role in their arid ecosystems. In this study, I focus on two species:

saguaro (*Carnegiea gigantea*) and organ pipe (*Stenocereus thurberi*). They contribute significantly to the ecosystem's dynamics, particularly in terms of ecohydrological processes. Monitoring the water storage, productivity, and distribution of columnar cacti populations is crucial for assessing the health of aridland ecosystems.

Figure 1.2: Scientists standing next to a large Mexican giant cardon (*Pachycereus pringlei*) during their data-collection expedition for the work discussed in section 1.1. Note that this cactus is a different species than the ones focused on in this work, but it still shares a very similar column-like structure and ecological role with those are (saguaro and organ pipe).

1.2.3 Point Clouds

Point clouds are a collection of data points with coordinates in 3D space. These points collectively depict the external surfaces of objects within the space and are commonly used to create highly accurate 3D models of real-world environments. Each point in a point cloud represents a single sample of an object's surface geometry captured by devices like LiDAR (light detection and ranging) scanners, or photogrammetry methods. LiDAR is a remote sensing method that uses light in the form of a pulsed laser to measure distances from the scanner to the solid objects in an environment, producing precise, three-dimensional information about the shape and surface characteristics of an area. Photogrammetry is a technique that involves capturing multiple photographs of a scene from different angles, and then using software to analyze and stitch together the images to determine the three-dimensional structure of the scene. In both cases, the point clouds can be scaled so that the resulting digital point cloud contains true-to-life distances, and can thus be used to measure the actual volume and surface area of objects contained in the point cloud. This study's dataset was gathered via photogrammetric methods, in which a drone or a cell phone was used for image capture depending on the size and obfuscation of the cactus specimen. Figure 1.3 shows one of the point clouds used.

Figure 1.3: Example saguaro with one image captured from the drone (left), the resulting 3D photogrammetric model (middle), and the isolated cactus body used for surface area and volume estimation (right).

1.2.4 Neural Networks

Neural networks are advanced machine learning models inspired by the brain, designed to recognize patterns and solve complex problems. At their core, neural networks consist of layers: an input layer that receives data, one or more hidden layers that process the data, and an output layer that delivers the final prediction. Each layer is made up of nodes, or neurons, which are connected to some number of nodes in the preceding and succeeding layers (see Figure 1.4 for a visulization). These connections have varying strengths, or weights. The fundamental operation in a node involves taking the weighted sum of its input connections, adding a bias term, and then typically applying a nonlinear function to this sum. This output is then passed on to the next layer. The entire network operates by moving data forward from the input to the output layer, a process known as feedforward.

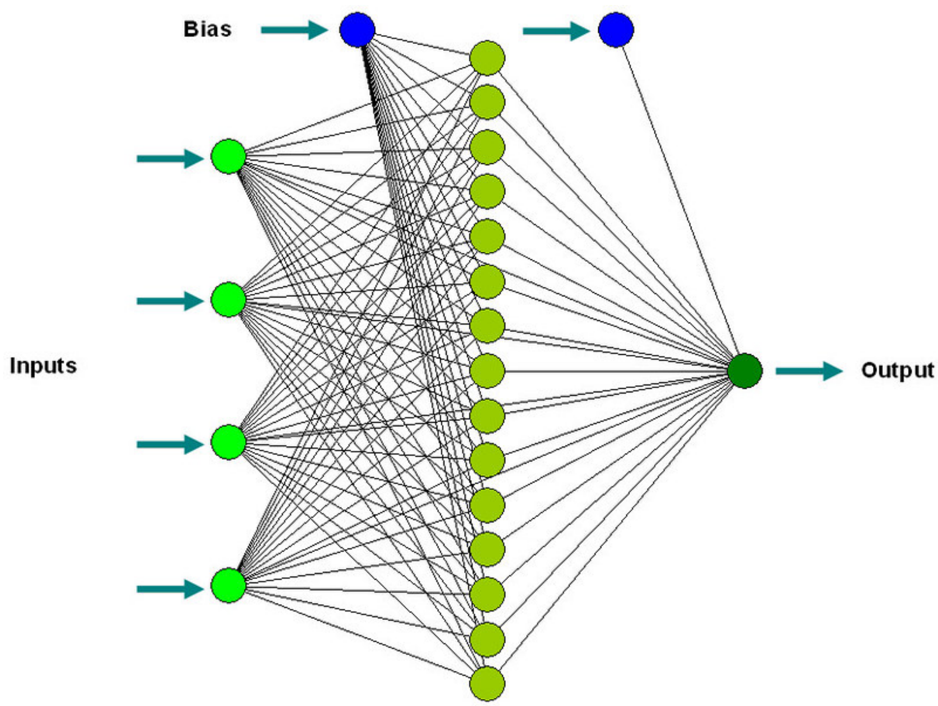

Figure 1.4: This figure demonstrates the basic form of a neural network. It shows an input layer, one hidden layer which does the data processing, and an output layer which makes a prediction. Each connection between layers in the figure will have a different weight value attached to it, and the bias term for each node is shown here as a separate node that is connected to all the nodes in its layer. So, each node's bias term can be thought of as an additional weighted connection to a bias node. Importantly, this diagram represents a simplified model; in practice, architectures typically involve multiple such hidden layers stacked sequentially, which is referred to as deep learning. This layering profoundly increases the network's capacity to extract and leverage intricate features from the data, embodying the core principles of deep learning for more sophisticated and accurate predictions.

However, for a neural network to make accurate predictions, it must be trained. Training involves adjusting the weights and biases from their initial random values to minimize the loss function, which is the neural network's way of measuring its error. It works by measuring the difference between the predicted outputs and the actual target values. This adjustment is achieved through a method known as backpropagation. During backpropagation, the contribution of each weight and bias to the overall error is calculated, starting from the output and moving backward to the input layer. The actual updates to weights and biases are performed using an optimization method called gradient descent. This method iteratively adjusts parameters to reduce the loss, with the magnitude of each update governed by the learning rate — a parameter that determines how big a step is taken at each iteration. Choosing the right learning rate is crucial; too large a rate can cause the model to oscillate around the minimum, while too small a rate can slow down the training process excessively.

Additionally, the training process involves deciding on a batch size, which is the number of data points the network processes before updating the weights and biases via backpropagation. When every sample in the training dataset has been used once for updating the model's weights and biases, this marks the completion of what is called an epoch. Training the model for multiple epochs gives it more time to tune its weights and biases to decrease the loss function. However, there is some point in which letting the model train for further epochs will not yield any performance improvements, and instead may harm the model by causing it to overfit to the training data.

The ability neural networks have to adapt and learn complex patterns is particularly suited for performing binary classification on point cloud data to identify columnar cacti. The unique challenges of interpreting point cloud data, such as varying densities and noise levels, demand a model that can adjust its internal parameters (weights and biases) through rigorous training cycles to discern subtle differences in 3D object forms. In this work, I perform experiments with two different kinds of neural network architectures: a multilayer perceptron (MLP) and a one-dimensional convolutional neural network (1D CNN). An MLP architecture is made up of fully-connected layers of nodes, as seen in Figure 1.4, whereas a 1D CNN performs an operation called convolution on the features of the input data. This convolution operation is meant to capture the relationships between data points and help the model learn local patterns in the data, whereas the MLP model analyzes each data point individually.

1.3 Objective

Accurately quantifying cacti, particularly on a landscape-scale, poses considerable challenges. Traditional methods of manual identification and volume calculation are labor-intensive, time-consuming, and often prone to inaccuracies due to the complex three-dimensional structures of cacti and their natural variations. Consequently, there arises a pressing need for automated approaches to detect and classify cacti within point cloud data, enabling more efficient and accurate volume calculations. Machine learning, particularly deep learning techniques, offers promising avenues for addressing the complexities of point cloud analysis. Therefore, the primary objective of this thesis is to develop a robust neural network model using the point cloud data previously gathered by S. E. Albeke, D. G. Williams, and K. R. Hultine, specifically tailored for the detection of columnar cacti within the data. The hypothesis this thesis seeks to test is whether an MLP or 1D CNN architecture is better suited for the task of extracting columnar cacti from point clouds via binary classification. With the goal of integrating the detected cactus bodies into volumetric analysis pipelines, this work aims to facilitate more efficient and reliable estimation of cactus biomass, water storage, productivity, and spatial distribution, so that we may enhance our understanding of aridland ecosystems and inform sustainable management practices.

Chapter 2

Literature Review

Recent advancements in point cloud processing have enabled incredibly high-fidelity three-dimensional models of physical spaces. With these data-rich models comes a need to efficiently analyze them through means such as segmentation, or the act of isolating specific objects. Though initial attempts at automated point cloud analysis were done without the application of deep learning techniques, the onset of neural network models has emerged as a powerful tool for classifying and segmenting point cloud data, offering promising accuracy and efficiency in recognizing complex objects within large datasets.

The foundation for current methodologies was laid by the introduction of the Point Cloud Library (PCL) in 2012, a pivotal resource that has facilitated the development and implementation of numerous point cloud processing algorithms [1]. Subsequent research focused on enhancing 3D object recognition through the integration of local and global features, leveraging the capabilities of PCL for improved accuracy and robustness in object classification, all without the aid of deep learning techniques [2]. By 2015, methodologies began to evolve, with studies demonstrating the efficacy of combining deep learning with traditional algorithms for fast and robust multi-view 3D object recognition [3]. This period marked a significant shift towards deep learning-centric approaches, culminating in innovative models like PointNet [4] and its successor, PointNet++ [5]. These networks introduced deep hierarchical feature learning directly on point sets, as opposed to transmuting the 3D point cloud data structure into a series of 2D projected images to subsequently input to the neural network, as done in [3]. This more direct approach is now the standard for point cloud classification and segmentation.

Direct input to a neural network is far from the only way to analyze point clouds via deep learning. A comparative study of point cloud classification highlighted the effectiveness of both point-based methods, such as PointNet, and neighborhood-based methods, such as a Dynamic Graph Convolutional Neural Network (DGCNN). This study elucidates the strengths and limitations of each approach, noting that while the DGCNN offers slightly higher overall accuracy with more parameters, PointNet achieves faster training and inference times, indicating a trade-off between performance and efficiency [6].

A series of comprehensive review and survey studies have provided a deeper insight into the advancements and applications of deep learning in 3D point cloud processing. Bello et. al. [7] provided an extensive review of deep learning techniques for raw point cloud data, emphasizing the importance of modeling local regions and the correlation between points within these regions for enhanced performance. Moreover, Wang et. al. [8] discussed a deep learning framework for large-scale 3D point cloud analysis and classification, underscoring the potential of combining traditional feature extraction methods with deep learning techniques for higher classification accuracy. Guo et. al. [9], and Hazer et. al. [10] presented surveys on numerous deep learning techniques on point clouds, highlighting the diverse approaches and their applicability to real-world tasks. By categorizing methods into convolution-based, point-wise multilayer perceptron (MLP), graph-based, and hierarchical data structure-based approaches, Hazer et. al. [10] sheds light on the technical properties, strengths, and limitations of these methods. This comparative study underscores the diversity of computational strategies available and aids in identifying the most effective techniques for specific research focuses. When performing a direct comparison of 3D point cloud object classification techniques from the known literature at the time, it was found that PointMLP [11] had the highest overall accuracy score on the popular ModelNet40 and ScanObjectNN datasets [10].

The continuous refinement of neural network architectures and training strategies has led to significant improvements in model performance and efficiency. For example, PointBLS [12] integrates deep learning with a broad learning system for enhanced point cloud classification, whereas PointNeXt [13] revisits PointNet++ with advanced training and scaling strategies, pushing the boundaries of what is achievable with point cloud analysis.

These developments have already found applications in ecological contexts, for instance, in classifying riverine species from UAV-derived photogrammetric point clouds [14], in segmenting trees from highly dense LiDAR point clouds [15], and in extracting tree crowns [16] and columnar cacti [17] from aerial imagery. Such research underscores the versatility and potential of neural networks in environmental applications, as the deep learning networks were able to effectively identify irregular, organic structures.

The trajectory of research in point cloud classification follows a trend towards more sophisticated and application-specific neural network models. By leveraging insights from comprehensive reviews and recent innovations in large-scale point cloud analysis, this research proposes a novel approach for automatically extracting columnar cacti from photogrammetric point clouds.