# Shape Based Classification and Segmentation Of 3D Point Clouds using Deep Learning

D. Jyothsna [1*, 2] and G. Ramesh Chandra [2]

[1] Department of Computer Science & Engineering, JNTUH, Hyderabad, Telangana, India
[2] Department of Computer Science & Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.

jyothsna.datti@gmail.com, rameshchandra_g@vnrvjiet.in

**Abstract.** The recognition of 3D objects that are represented as point cloud has become a very vital topic for research in the field of Computer Science. The processing of 3D point clouds has procured a lot of attention from researchers in the modern era. However, as point clouds have a complex representation, it is still a challenging task to capture 3D objects using LiDAR (Light Detection and Ranging) devices. The goal of our work is to identify an object that is represented as a point cloud accurately using 3D Deep Learning. The scope of the project is very vast, and its applications include AR/VR (augmented/virtual reality), self- driving vehicles, robotics, etc. The work aims to classify the 3D objects and also to identify various integrated parts of the object by performing segmentation by accessing the 3d point clouds directly on modelnet10 and shapenet for classification and segmentation respectively. In classification, the captured object is classified into several classes such as a chair, table, nightstand, sofa, bed, etc. Using the most suitable deep learning model we perform classification on the point clouds obtained from the ModelNet10 dataset and segment the objects into parts using shapenet dataset. Also, experimentation was done on different learning rate parameters for better results.

**Keywords:** 3D point cloud, Classification, Segmentation, Deep Learning.

## 1 Introduction

When it comes to creating robot vehicles, self-navigating robots, and a variety of other real-world applications, environmental awareness is crucial. There are a variety of sensors that can capture a 360-degree view of real-world things. Some of them are LiDAR, RADAR, cameras, etc. LiDAR (Light detection and ranging) works on the principle of Time of Flight [15]. The distance between the object and sensor is calculated based on the time taken to capture the reflected light ray by the receiver present in the LiDAR device [9] [15]. Previously the 3D data produced by the LIDAR device was most useful for providing accurate geospatial measurements.

The point cloud [16] is one of the most used representation forms of a 3D object. However, as point clouds have a difficult structure, it is difficult to classify them. While


C. Kiran Mai et al. (eds.), *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023)*, Atlantis Highlights in Computer Sciences 18,
https://doi.org/10.2991/978-94-6463-314-6_6

giving point cloud data as input to a deep learning neural network, frequently the data is converted into other representations like voxel grids, meshes, images, etc. due to their irregular format. The developed architecture performs classification and per point segmentation to produce class labels of integrated objects in a given input point cloud. In this paper, related works are covered under section 2, details of datasets used in section 3, the proposed methodology in section 4, results in section 5, and conclusion and future scope in section 6.

## 2 Related Work

The data of 3 Dimensions can be represented in various formats like meshes, point clouds, depth images, volumetric grids, multi view 2D images, projections, CAD, and graph etc. [10,11]. In the existing systems (traditional architectures) the point cloud is converted to other representations like voxel or grids since point clouds are in irregular format before giving them as an input to the network. But in the application of this transformation on the data will make it more voluminous, needs more memory and time to generate surfaces and boundaries from point clouds. Deep learning on 3D point clouds, on the other hand, faces several important problems, including short dataset sizes, unstructured and highly dimensional in nature. So, our aim is to focus on simple representation for 3D geometry using point clouds.

In MVCNN [12] uses multi-view features obtained from different2D views which are taken from single 3D object. This network provides CNN to recognize the shape which considers each 2D image independently to show that 3D shape can be recognized using single 2d view [8]. However, the network also provides the global feature obtained by max pooling but max pooling considers only important features leading to information loss. 3D shape understanding is discussed in [3] with unsupervised learning approach. [6] used graphs to classify the point clouds using sampling [7]. With feature description map, [5] proposed a model to classify objects in large scenes. In [15], a hybrid model is proposed class labels point by point.

In [2] presented a transformer based network suitable for unstructured, disordered point cloud data, named PCT for point cloud processing. It is based on the transformer that has achieved major success in NLP. It is invariant to permutations, so it is suitable for processing point clouds. Pure Segmentation algorithms need strict geometric constraints and rules. In [1] proposed a fully unsupervised region growing segmentation process for effective clustering of 3D point clouds by using self-learning heuristic approach on the S3DIS dataset. But this approach is not suitable for complex objects. [4] worked on residual networks and classified objects using residual fusion network.

## 3 Dataset

A huge number of datasets are available for evaluating the deep learning algorithms. ModelNet [13], ShapeNet [14], ScanNet, Semantic3D, and KITTI are among the famous datasets used in classification and segmentation of 3D point clouds.

### 3.1 ModelNet

For classification, the proposed system uses ModelNet10 dataset which is subset of ModelNet40 [13] dataset. ModelNet40 consists of 12,311 shapes from 40 different classes, which are divided into 9,843 smaples (80%) for training dataset and 2,468 samples (20%) for testing dataset. ModelNet10 is a subset of ModelNet40, containing 4,899 shapes from 10 different categories. There are 80% i.e., 3,991 shapes for training dataset and 20% i.e. 908 shapes for testing dataset. The CAD Models are in Object File Format (OFF). Fig. 1 shows some of the sample shapes of 5 different classes from modelnet 10.

### 3.2 ShapeNet

ShapeNet [14] is a major repository for 3D CAD models developed by researchers from various Universities. This repository consists of over 300M CAD models with 220,000 categorized into 3,135 classes. ShapeNet Parts subset consists of 31,693 meshes classified into 16 familiar object classes such as a table, chair, plane, etc. Each shape ground truth contains 2-5 parts (with a total of 50 part classes) [14]. ShapeNetCore is a subset of the ShapeNet dataset that provides single clean 3D models. It has roughly 51,300 distinct 3D models that cover 55 popular object categories divided into 12 classes [14]. Refer Fig.1 for the samples of shapenet dataset.


**Fig. 1.** Sample Images of Modelnet10 and Shapenet [Source: https:// link.springer.com/article /10.1007/s11263-022-01650-4, https://arxiv.org/pdf/2004.09411.pdf]

## 4 Methodology

The proposed system uses Point Net Architecture which consumes point clouds directly without any transformation. Point Net Model treats each point using independent MLPs (Multi-layer perceptron) and then combines the global features using an aggregate function. There are mainly 2 modules: Classification and Segmentation.

Classification involves predicting and classifying an object to which class it is belongs. Methods for this module typically study the embedding of each factor first after which extract global features embedding from the point cloud using concatenation method. Segmentation means division into separate sections or clustering of parts that belong to the same class. It helps us to define the boundaries of different sections of the same object. Understanding of the global geometric features and local details of every single point is very much needed for the 3D point cloud segmentation.

Based on the segmentation level and scale, the 3D point cloud segmentation methods are classified as semantic segmentation (scene level), instance segmentation (object level), part segmentation (part level). The Fig. 2 describes the overview of architecture.

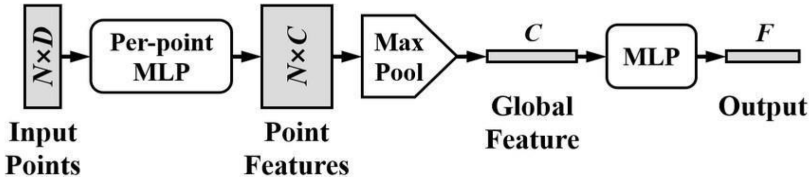


**Fig. 2.** System Architecture [Source: https://ieeexplore.ieee.org/document/8578976]

Below Fig. 3 describes the architecture for classification and segmentation using point net. It takes n points each of 3 dimension (n*3) applies input transformation and feature transformation using T- net [2] a transformer network that makes point clouds invariant to permutation. The n*3 input points are fed to multiple MLPs and each MLP outputs a higher order input point, and then aggregates point features to obtain global features by max pooling.

Segmentation can be performed by extending the classification network. Local features (the output generated after performing the second transformation network (t-net) and global features (output obtained from max pooling) are combined for each point. The combined features of each point are fed to MLP layer. The output is segmentation scores for N classes with M parts (N*M). For the optimization purpose, Adam optimizer is used with the learning rate of 0.001 in both classification and segmentation tasks. Also, experimented on different learning rates 0.001, 0.002 and 0.003.

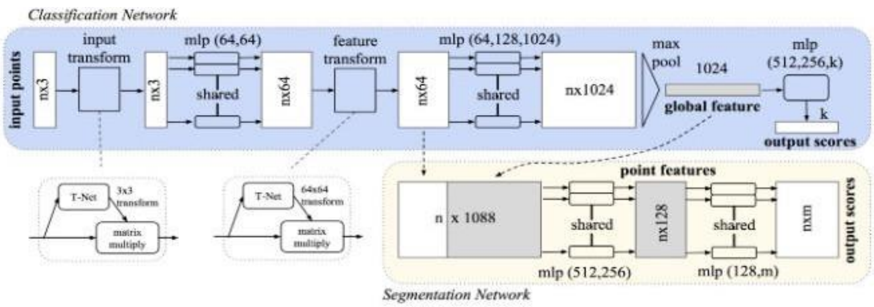


**Fig. 3.** Detailed Architecture for Classification and Segmentation [Source: https://arxiv.org/pdf/1612.00593.pdf]

## 5 Results and Discussions

The aim the project is to develop a model (Deep Net) that directly processes the 3D point clouds and performs classification and segmentation tasks. The model is trained on Modelnet10 dataset for classification and shapenet dataset for segmentation. Experiments were done on each epoch to find the losses for the learning rate parameters 0.01, 0.02 and 0.03, however learning rate 0.01 yielded better results with low loss values. Refer Fig. 6 to verify the loss values.

### 5.1 Classification

This model takes 1024 input points with a batch size of 32 and the no. of epochs is 60. The training dataset is 3991 shapes, validation dataset is 908 shapes and the no of classes is 10 .The model takes a point cloud that has 1024 points as an input and performs input transformation using a T-Net to make it invariant to spatial transformations. Then MLP is applied for processing input and finally, Max Pooling operation is performed to extract the global feature of the object present in the input point cloud. The following Fig. 4 represents the results produced by the model after performing classification.

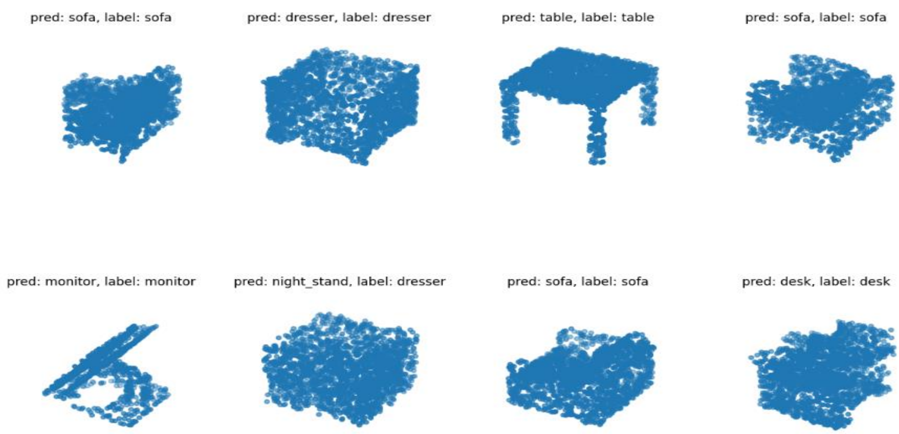


**Fig. 4.** Classification Results of Point Cloud

To evaluate the correctness of the prediction results, confusion matrix is generated for all the 10 classes between true labels and the predicted labels. The highest accuracy resulted with the object ‘chair’ that detected all true labels for prediction, Followed by the object ‘bed’ with 96%, sofa with 93% true predictions etc. The overall accuracy is 81%. Refer Table 1 and Fig. 5 to find the prediction accuracy of each class in modelnet 10 dataset.

**Table 1.** Accuracy of individual objects in Modelnet10dataset

| Class | Prediction Accuracy (%) |
|---|---|
| **BathTub** | 69 |
| **Bed** | 96 |
| **Chair** | 100 |
| **Desk** | 59 |
| **Dresser** | 66 |
| **Monitor** | 83 |
| **Night Stand** | 80 |
| **Sofa** | 93 |
| **Table** | 94 |
| **Toilet** | 70 |

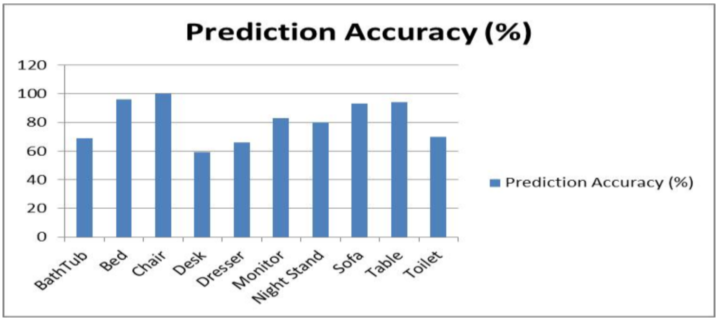


**Fig. 5.** Prediction Accuracy of the classes in ModelNet10

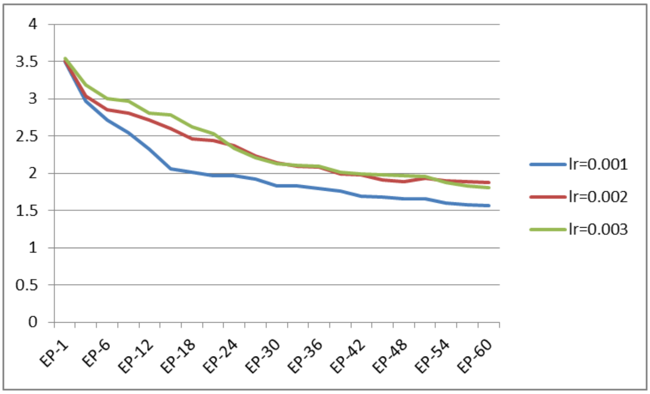


**Fig. 6.** Losses for different epochs for the learning rate parameter 0.01, 0.02 and 0.03

### 5.2 Segmentation

The model is trained with the 1024 points with a batch size of 32 and the no. of epochs is 60. The model consider both the input points that have been acquired after performing feature transformation and the global feature acquired by performing max pooling. The model processes them and produces output scores. Then the segmentation results can be produced based on the output scores produced by the model. The following Fig. 7 represents the results produced by the model after performing Segmentation. The blue color represents wings of the plane, green color indicates body of the plane, red color represents the tail of the plane and pink color represents the engine of the plane. The highest accuracy of the segmentation is 85%.

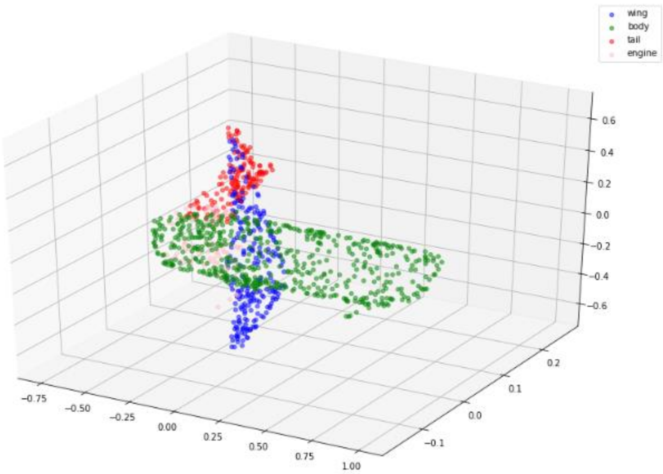


**Fig. 7.** Segmentation Results of Airplane

## 6 Conclusions and Future Scope

The primary aim of this work is to classify and segment the point cloud shapes using pointnet architecture on Modelnet 10 and shapenet datasets respectively. The given input shape is successfully classified into classes like table, chair, sofa, monitor, toilet, bathtub, etc. and segmented with an accuracy of 85%. To achieve these results, experimentation was done on one of the hyper parameters namely learning rate at three different values to compute the results with low losses. Hence the aim of the project is fulfilled by processing the 3D point clouds directly without converting them into other representation forms like voxel grids, meshes, 2D images, etc. The developed architecture can perform classification and segmentation with good accuracy. This project can further be developed to learn deep point set features efficiently and robustly. The model can also be extended by classifying the text that is present on the objects and by improving its accuracy. Also, the model can be used as the backbone for feature extraction for object detection in 3d point clouds.

## Acknowledgement:

This research work was supported by the short range LIDAR project funded by the Directorate of Extramural Research & Intellectual Property Rights (ER&IPR), Defence Research & Development Organisation (DRDO) grant no. ERIP/ER/202108001/M/01/1786.

## References

1. Poux, F., et al, "Unsupervised Segmentation of Indoor 3d Point Cloud: Application to Object-based Classification" Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., https://doi.org/10.5194/isprs-archives-XLIV-4-W1-2020-111-2020, 2020.
2. Mang-HaoGuo et al, "PCT: Point Cloud Transformer" Computational Visual Media,202, Vol. 7, No. 2, Pages:187 – 199, https://doi.org/10.1007/s41095-021-0229-5, June 2021.
3. Matheus Gadelha, et. Al,"Multiresolution Tree Networks for 3D Point Cloud Processing" , https://doi.org/10.1007/978-3-030-01234-2_7, ECCV 2018.
4. C. -H. Chiang, C. -H. Kuo, C. -C. Lin and H. -T. Chiang, "3D Point Cloud Classification for Autonomous Driving via Dense-Residual Fusion Network," in IEEE Access, vol. 8, pp. 163775-163783, 2020, doi: 10.1109/ACCESS.2020.3021191.
5. Wang, Lei et. Al, "Large-scale 3D Point Cloud Classification Based On Feature Description Matrix By CNN". CASA 2018: Proceedings of the 31st International Conference on Computer Animation and Social Agents. 43-47. 10.1145/3205326.3205355.
6. Zhang, Yingxue & Rabbat, Michael. (2018). A Graph-CNN for 3D Point Cloud Classification. 6279-6283. 10.1109/ICASSP.2018.8462291.
7. Lang Itai et al, "SampleNet: Differentiable Point Cloud Sampling", https://doi.org/10.48550/arXiv.1912.03663 , Apr 2020.
8. Su, Hang et Al, Erik. (2015). "Multi-view Convolutional Neural Networks for 3D Shape Recognition". 10.1109/ICCV.2015.114.
9. C. Premebida, et Al, "Exploiting LIDAR-based features on pedestrian detection in urban scenarios," 2009 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 2009, pp. 1-6, doi: 10.1109/ITSC.2009.5309697.
10. Guo, Yulan & Wang, Hanyun & Hu, Qingyong & Liu, Hao & Liu, Li & Bennamoun, Mohammed. (2020). Deep Learning for 3D Point Clouds: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2020.3005434.
11. Xinliang Zhang et al, "Hybrid featureCNN model for point cloud classification and segmentation" 21 Feb 2021, https://doi.org/10.1049/iet-ipr.2020.0658.
12. Su, Hang & Maji, Subhransu , et al, (2015). Multi-view Convolutional Neural Networks for 3D Shape Recognition. 10.1109/ICCV.2015.114.
13. ModelNet Dataset, "https://modelnet.cs.princeton.edu".
14. ShapeNet Dataset, https://shapenet.org.
15. Lidar " https://en.wikipedia.org/wiki/Lidar".
16. Point cloud https://en.wikipedia.org/wiki/point cloud.