 ISO/IEC JTC 1/SC 29/WG 7 N234

**ISO/IEC JTC 1/SC 29/WG 7  
MPEG 3D Graphics Coding   
Convenorship: AFNOR (France)**

**Document type:** Output Document

**Title:** Preliminary Dataset Collection for AI-based Point Cloud Experiments

**Status:** Approved

**Date of document:** 2021-12-02

**Source:** ISO/IEC JTC 1/SC 29/WG 7

**Expected action:** None

**Action due date:** None

**No. of pages:** 6 (with cover page)

**Email of Convenor:** marius.preda @ imt . fr

**Committee URL:** [https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg7](https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3)

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 7 MPEG 3D Graphics Coding**

**ISO/IEC JTC 1/SC 29/WG 7 N234**

**Oct 2021, Virtual**

|  |  |
| --- | --- |
| **Title** | **Preliminary Dataset Collection for AI-based Point Cloud Experiments** |
| **Source** | **WG7, MPEG 3D Graphics Coding** |
| **Status** | **Approved** |
| **Serial Number** | **21007** |

Abstract

An exploration AhG on “AI-based coding for graphics” was established during the 4th WG7 /the 135th MPEG) meeting in July 2021. This document provides a description of some of the available opensource point cloud datasets and their applicable terms of use.

# Introduction

One of the most important aspects of learning-based systems is the dataset that is used by a learning-based method to learn the task. In this document we discuss some existing open-source datasets that can be considered within the MPEG group for the purpose of learning-based point cloud compression and analysis.

# Proposed Datasets terms of use

We propose the following four datasets from the plethora of existing open-source datasets: ShapeNetCore dataset, ModelNet40 dataset, SemanticKitti dataset, and Ford sequences from GPCC CTC. The first two datasets, ShapeNetCore and ModelNet, are CAD models of individual objects whereas Kitti and Ford are LiDAR scans of outdoor scenes. As highlighted in the MPEG input document [1], these datasets can be put in the following categories based on the type of data and point density: *Dense point clouds* for ModelNet40 and ShapeNetCore; and *Automotive Frame point clouds* for Kitti and Ford.

We now provide some usage conditions specific to each dataset below. Ford sequences are already part of MPEG CTC and are thus skipped.

## ShapeNet

ShapeNet terms of use can be found at [2] :

* Two out of the seven terms and conditions are highlighted below:
  + Researcher shall use the Database only for non-commercial research and educational purposes.
  + If Researcher is employed by a for-profit, commercial entity, Researcher's employer shall also be bound by these terms and conditions, and Researcher hereby represents that he or she is fully authorized to enter into this agreement on behalf of such employer.
* The group needs to discuss if these usage conditions allow usage of this data for training and evaluation in regard to learning-based methods.

## ModelNet

ModelNet copyright can be found at [3] and is reproduced below:

* “All CAD models are downloaded from the Internet and the original authors hold the copyright of the CAD models. The label of the data was obtained by us via Amazon Mechanical Turk service and it is provided freely. This dataset is provided for the convenience of academic research only.”
* The group needs to evaluate if this copyright allows usage of this data for training and evaluation in regard to learning-based methods.

## Kitti, SemanticKitti

Kitti dataset copyright can be found in [4]:

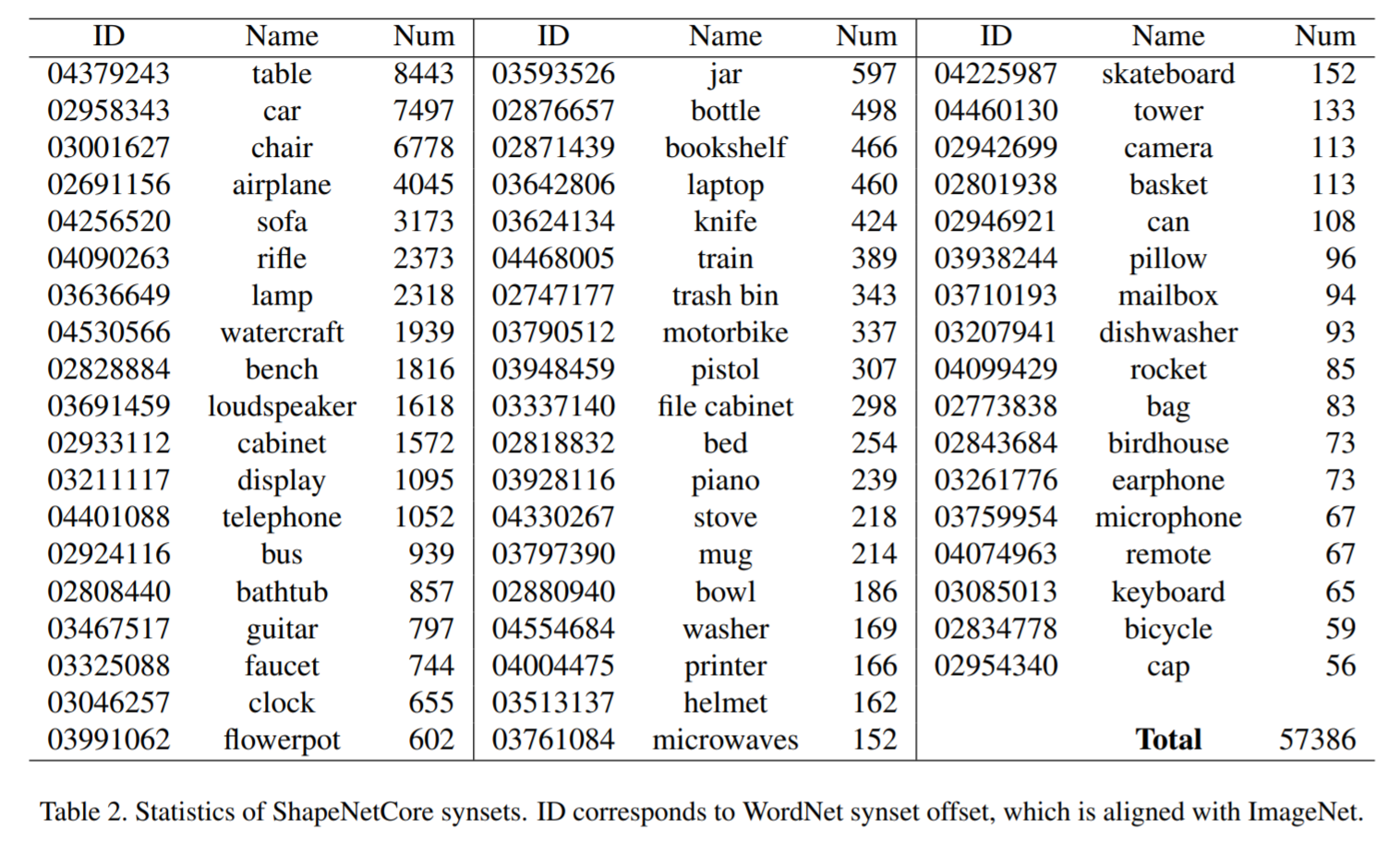
* “All datasets and benchmarks on this page are copyright by us and published under the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0](http://creativecommons.org/licenses/by-nc-sa/3.0/) License. This means that you must attribute the work in the manner specified by the authors, you may not use this work for commercial purposes and if you alter, transform, or build upon this work, you may distribute the resulting work only under the same license.”
* The group needs to evaluate if this license allows usage of this data for training and evaluation in regard to learning-based methods.

# Dense point cloud dataset characteristics

Dense point cloud dataset composed of single object can be used for training networks designed to: a) compress this object level data, and b) perform vision tasks such as object recognition, object segmentation, object part segmentation, etc.

## ShapeNetCore Dataset

ShapeNet is an ongoing effort to establish a richly-annotated, large-scale dataset of 3D shapes. ShapeNetCore dataset, a subset of the full ShapeNet dataset, contains 51,300 3D objects models from 55 common object categories. The data is provided as 3D meshes (dense connected graphs with faces) which can be sampled to obtain point clouds using several tools available on the ShapeNet website. PyTorch3D now also supports easy loading of ShapeNetCore dataset for that purpose. Example objects in ShapeNetCore are highlighted in the following table taken from [5].



## ModelNet dataset

ModelNet is a large 3D CAD model dataset containing various shapes with 151,128 models in 660 unique object categories. Two smaller subsets, ModelNet10 and ModelNet40 are also available each containing 10 and 40 unique object categories. The CAD models are in the format of “OFF” (object file format) files, which are mesh files from which point clouds can be sampled with required sampling density. Below is a word cloud representing the various categories in the original ModelNet dataset.



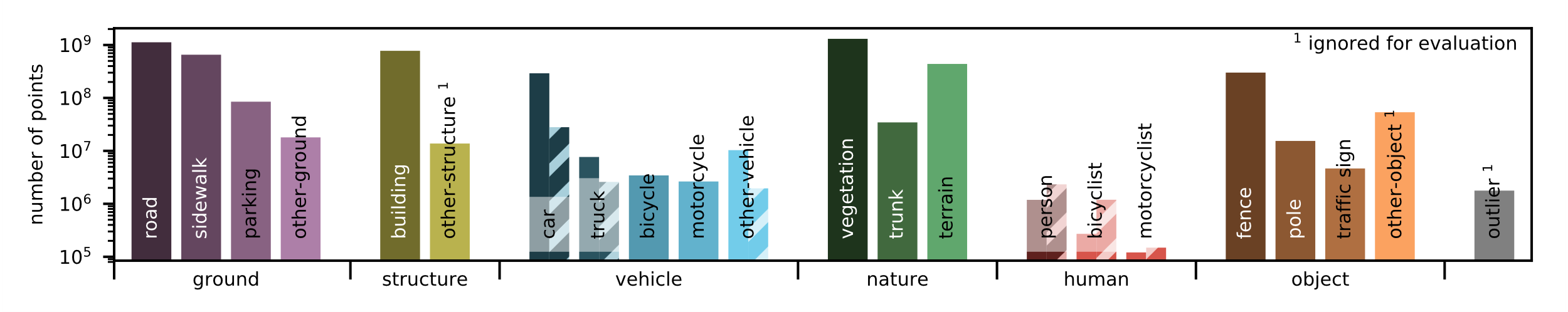
# Automotive Frame point cloud dataset characteristics

Automotive Frame point clouds are LiDAR point clouds obtained from sensors mounted on automotive driving outdoors. These point clouds consist of a sequence of frames obtained in a series with a sensor specified fps (frames per second) parameter. This kind of data can be used to train networks to: a) compress LiDAR data, and b) perform specific machine tasks such as object recognition, object segmentation, motion tracking, motion segmentation, semantic segmentation, etc.

## Kitti Dataset

Kitti dataset is a large-scale dataset with LiDAR scans obtained in outdoor self-driving settings containing large amount of raw LiDAR data, along with other data modalities (e.g., RGB) and some labels. SemanticKitti developed by a group of researchers from University of Bonn, contains dense annotations for all the sequences in the odometry task (from within the original Kitti dataset). Out of 43552 scans in total from Kitti, SemanticKitti provides ground truth annotations under several categories for 23,311 scans for training purposes.

The data itself is provided as “.bin” files in continuous range (-96,96). However, with conversion to integer values the data lies within range (0,2^16). The average number of points in each scan is around ~120K.

Below is an example of the kind of dense (point-wise) annotations available in SemanticKitti.

## Ford sequences from GPCC CTC

Re-use of a refined current MPEG dataset for testing was proposed in [6]. We propose to use the three frame-based “Ford” sequences [7] which contain 18bit data with 1500 frames in each sequence. Moreover, each sequence has on average ~100K points per frame.

# Dataset partitioning

The point clouds data in the datasets should be divided into two categories:

* Benchmarking set: This set of data is entirely reserved for benchmarking the performance of the learning-based PCC codec. It should not be used for training.
* Trainval set: This set of data is used to train and validate the learning-based codec. It can be used freely.

# Proposed mandates to be explored

* The biggest question is for the group to evaluate the usage conditions of the datasets proposed in this document to infer if the proposed datasets can be used within the group.
* Should the AI methods be benchmarked on data that is from the same dataset as the trainval data or a different dataset, or both?
  + If a codec performs well on data from a different dataset than one used in trainval, it shows better generalizability.
* How much data should be used for training?
  + Networks with many parameters will require extensive data, while others can do with less data due to smaller number of parameters.
* Should one model be trained for all rate points or should separate models for each target rate point be trained?
* When evaluating various existing point cloud compression architectures, in [8], the authors decided to compress down-sampled point clouds instead of full resolution point clouds. This should still be a topic of discussion in the next meeting as compression of full resolution point clouds is an important issue.
* We also reiterate the already presented call within the group to gather newer and diverse datasets for training and evaluation of the AI methods.

# References

|  |  |
| --- | --- |
| [1] | K. M. D. Flynn, "G-PCC: Updating common test conditions," *MPEG,* no. m57468, Jul 2021. |
| [2] | "ShapeNet," [Online]. Available: https://shapenet.org/terms. |
| [3] | "ModelNet," [Online]. Available: https://modelnet.cs.princeton.edu/. |
| [4] | "Kitti VIsion Benchmark Suite," [Online]. Available: http://www.cvlibs.net/datasets/kitti/index.php. |
| [5] | Z. W. e. al., "3d shapenets: A deep representation for volumetric shapes," *IEEE CVPR,* 2015. |
| [6] | J. Pang, M. A. Lodhi, G. Martin-Cocher and D. Tian, *[PCC] AI-based Point Cloud Compression for New PCC,* Online: MPEG m56776, Apr 2021. |
| [7] | WG7, MPEG 3DG, *Common Test Conditions for G-PCC,* Online: MPEG N00106, Jul 2021. |
| [8] | A. Zaghetto, D. Graziosi and A. Tabatabai, *Discussion on Point Cloud Geometry Compression Using Machine Learning,* Online: MPEG m57301, Jul 2021. |