Learning based point cloud compression: A stability analysis

Joao Prazeres
joao.prazeres@ubi.pt
Rafael Rodrigues
rafael.rodrigues@ubi.pt
Manuela Pereira
m.pereira@di.ubi.pt
Antonio M. G. Pinheiro
pinheiro@ubi.pt

Universidade da Beira Interior & Instituto de Telecomunicacoes
Rua Marquês D'Ávila e Bolama
6200-001, Covilhã, PT

Abstract

In this paper, a study on the stability of three deep learning point cloud compression solutions, notably ADLPCC, PCC GEO CNNv2, and PCGCv2 is presented. This study aims to show one of the problems of deep learning based codecs, that results in instability on the codec performance. These behaviours appear in unusual cases. The same architecture with a new training will most likely solve the problem. An example of a point cloud representing cultural heritage is used in this paper for demonstration purposes. Three different training sessions were conducted using the default training set and cost function of each of the considered codecs. Across each epoch of the training sessions, the objective quality metric MSE PSNR D1 was computed. The final result of each training session was objectively to the default implementation of the codecs, using the MSE PSNR D1 and PCQM metrics.

1 Introduction

As a model for 3D data representation, point clouds are becoming increasingly popular. This method of representation uses a set of (x,y,z) Cartesian coordinates to represent 3D data, and each coordinate may have a number of attributes attached to it, including RGB components, reflective information, physical sensor information, or normal vectors. Point clouds can offer incredibly accurate depictions of a particular artifact, object, landscape, or building, creating an outstanding 3D representation model. Point clouds might, however, hold a huge amount of data, which is challenging to transmit, store, and manipulate. Therefore, efficient point cloud coding methods as well as accurate performance evaluation models are crucial for the success of this format.

2 Learning based Point Cloud Coding

Several machine learning-based point cloud coding methods have been recently proposed [1, 5, 7]. The problem is that the same model trained with the same data and rate distortion relation can, in some cases, result in completely different performance. Some cases even result in worse distortion (or quality) at higher bit rates. However, if trained again with exactly the same conditions, the model can lead to the usual behavior, which means a higher bit rate corresponds to lower distortion. Although this is an unusual behavior, there are cases where a given deep learning-based codec can lead to these unexpected and undesired instabilities. This research follows the model established in a previous study [4]. The model consists of analyzing the compression performance of deep-learning based codecs throughout three training sessions under identical conditions. Throughout the learning process, MSE PSNR D1 [6] is computed at each training epoch for the coded point cloud. The final result of each training session is then compared to the default trained codec to verify if there are relevant differences. For this evaluation, three deep-learning based codecs were used, notably PCGCv2 [7], PCC GEO CNN v2 [5], and ADLPCC [1].

3 Stability Analysis of Deep Learning-based Codecs

For the considered solutions, the global loss function depends on the distortion of the encoded point clouds and the encoding bit rate. The encoding bit rate is estimated differently for each codec. PCGCv2 estimates the distortion from the Binary Cross-entropy loss function (BCE), $BCE = -\frac{1}{N}\sum_i (x_i \log(p_i) + (1 - x_i \log(1 - p_i)))$, where x_i is the true binary occupancy value of voxel i, and p_i is its occupancy probability output given by the model.
PCC GEO CNN v2 and ADLPCC use a focal variation of the BCE to address imbalances between empty and occupied voxels in more sparse point clouds.

3.1 Training procedure

The codecs were trained three times using the default implementations. The parameters for each training session were kept exactly the same, in order to be able to directly compare the results. The global loss function used by the three tested codecs is based in a linear relation between the distortion D and Rate R ($J = \alpha D + \beta R$). In this experiment, PCGCv2 [7] α was set to $\{16, 4, 0.75\}$ while β was left at 1. PCC GEO CNN v2 trains individual models for each rate distortion tradeoff [5], with α=$\{3\times 10^{-4}, 10^{-4}, 5\times 10^{-5}, 2\times 10^{-5}, 10^{-5}\}$ and $\beta = 1$. Both PCGCv2 and PCC GEO CNN are initially trained for the higher bit rate, and then the lower bit rates are derived using the previous bit rate/distortion settings for the initial training.

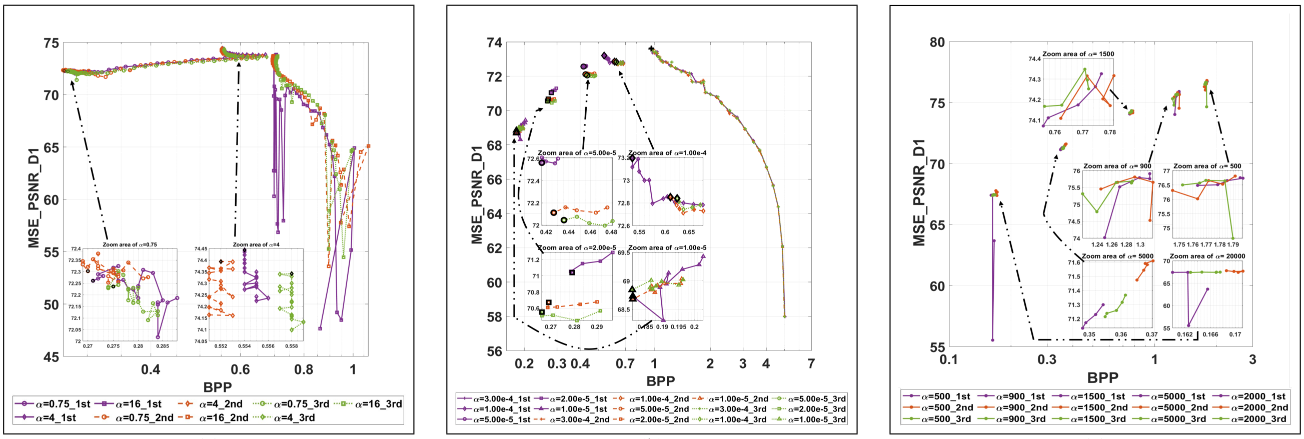
The global loss function of ADLPCC is given by $\beta = 1$ and the coding rate R is estimated during training as the summed entropy of its autoencoder and variational autoencoder latent representations. In order to obtain several rate distortion tradeoff points, different α values are considered, thus varying the weight of the rate. The model was trained with $\alpha = \{500, 900, 1500, 5000, 20000\}$. For each value of α, the codec was trained with the BCE focal loss function parameter
with values $\{0.5, 0.6, 0.7, 0.8, 0.9\}$[1].

3.2 Quality Evaluation

The objective quality metrics MSE PSNR D1 [6] and PCQM [2] were used to benchmark the training sessions. PCQM is a metric that uses geometry and color for quality evaluation and was selected for its ability to accurately represent subjective quality [3]. As it requires color, the reference point cloud color information was mapped onto the distorted geometry to create a colored point cloud. The resulting cloud was then encoded with G-PCC using the `lossless-geometry-lossy-atts` mode, so that the distorted geometry remains untouched.

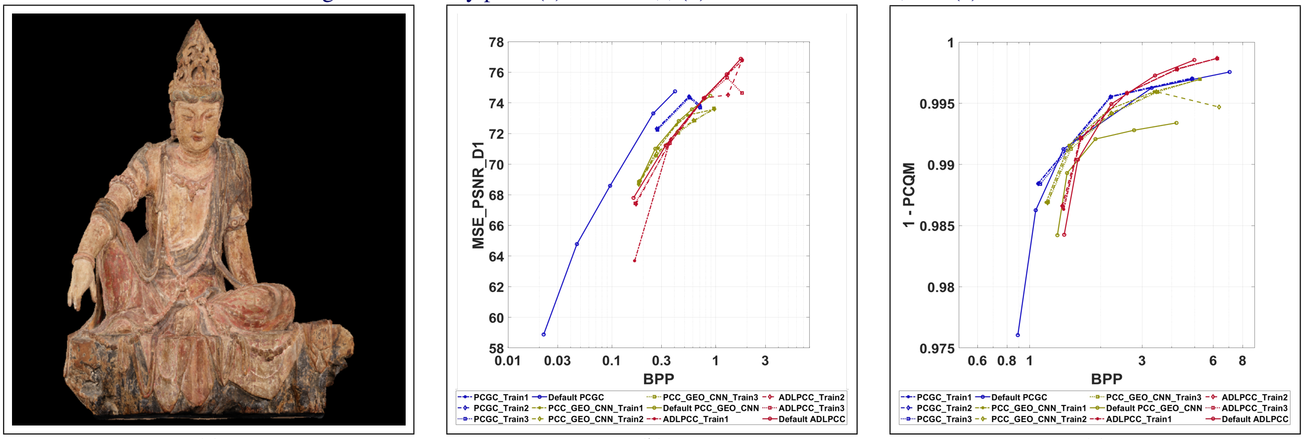
The MSE PSNR D1 plots across all training sessions for the three codecs are represented in Fig. 1(a), (b) and (c)) when the point cloud used as an example (*Guanyin* represented in Fig. 2(a)) is encoded using the codec that resulted in each training epoch. All codecs result in slightly different bit rate/distortion relations for each of the training sessions, which is also undesirable.

The MSE PSNR D1 results for PCGCv2 show a similar behaviour across three training sessions. However, cases of instability are found in the intermediate epochs for $\alpha = 16$, especially at the higher bit rates of the first training session. PCC GEO CNNv2 is fairly stable. At intermediate rates (1×10^{-4}, 5×10^{-5}, and 2×10^{-5}), the MSE PSNR D1 values across training sessions converge to a common working point. No situation where the final coding output depends heavily on the training session was found. ADLPCC is highly stable, with minimal fluctuation in encoding steps over epochs, especially when trained with $\alpha = 20000$. The first train shows a sudden drop in an intermediate epoch, and the third train shows a drop in the final epoch. The quality of the training process for other α values was consistent across bit rates.



(a) (b) (c)

Figure 1: Stability plots (a) PCGCv2, , (b) PCC GEO CNN v2, and (c): ADLPCC.



(a) (b) (c)

Figure 2: (a): *Guanyin* (b): MSE PSNR D1 plots and, (c): PCQM plots for each of the defined operating points for each codec.

Fig. 2(b) shows the bit rate distortion plot represented using MSE PSNR D1 and Fig. 2(c) the PCQM. It can be observed that both metrics reveal a different performance for each training session. Moreover, the training sessions with the best performance for this point cloud might not be the best for other point clouds, as was further observed in this study. This performance instability and dependency of the training session are highly undesirable for codecs.

ADLPCC presents the most stable results. The MSE PSNR D1 plot suggest that the models from the three training sessions perform similarly to the default codec. It is also shown that the codec has a degree of instability in the higher rates, as training session 2 produced a lower quality encoding, as well as a quality decrease in the highest bit rate for training session 3. The lowest bitrate also shows some levels of instability. The PCQM plots show that the models outperform the default codec at lower bit rates but fails to reach the quality levels of the default codec at higher bit rates. The PCC GEO CNNv2 results are slightly unstable. The MSE PSNR D1 results from the three training sessions show good convergence in the lower and higher bit rates, but the middle bit rates show some variations. The PCQM results are quite unstable. The instability in the medium bit rates is still present and is accentuated in the higher bit rates, especially in training session 2, where quality significantly drops. The PCQM plots show that the models exceed the default codec by a significant margin further accentuating the quality difference at high bit rates. PGCCv2 is the most stable codec. It achieved similar results across the training sessions. Nonetheless, it should be noted that all training sessions shows a very undesirable rate drop at the highest bit rate. As the authors do not specify their lambda trade-off in the current implementation, outcomes may vary depending on the default codec and training sessions.

4 Conclusions

This study showed that the most well-known deep learning-based coding solutions for point clouds reveal a performance with some degree of dependency on the training, independently of keeping all training conditions (training data and loss functions) unchanged. Although in most cases the level of stability is high, the fact that some cases appear with instability in the bit rate/distortion might reveal problems for the general adoption of these coding solutions. The instability problems evident during the training epochs might cause a codec to result in very bad performance for a given content, in contrast to other content where a high level of performance is reached. Furthermore, the higher instability is likely to appear at higher bit rates, where the quality is expected to be better.

References

[1] André FR Guarda, Nuno MM Rodrigues, and Fernando Pereira. Adaptive deep learning-based point cloud geometry coding. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):415–430, 2020.

[2] Gabriel Meynet et al. PCQM: A full-reference quality metric for colored 3D point clouds. In *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2020.

[3] Joao Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio MG Pinheiro. Quality evaluation of machine learning-based point cloud coding solutions. In *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis*, pages 57–65, 2022.

[4] João Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio M. G. Pinheiro. On the stability of point cloud machine learning based coding. In *10th European Workshop on Visual Information Processing (EUVIP)*, pages 1–6, 2022.

[5] Maurice Quach, Giuseppe Valenzise, and Frederic Dufaux. Improved deep point cloud geometry compression. In *IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2020.

[6] Dong Tian, Hideaki Ochimizu, Chen Feng, Robert Cohen, and Anthony Vetro. Geometric distortion metrics for point cloud compression. In *ICIP*, 2017.

[7] Jianqiang Wang, Dandan Ding, Zhu Li, and Zhan Ma. Multiscale point cloud geometry compression. In *Data Compression Conference (DCC)*, pages 73–82. IEEE, 2021.