A Video Dataset for Learning-based Visual Data Compression and Analysis

Xiaozhong Xu  
Tencent Media Lab  
Palo Alto, CA 94306, USA  
xiaozhongxu@tencent.com

Shan Liu  
Tencent Media Lab  
Palo Alto, CA 94306, USA  
shanl@tencent.com

Zeqiang Li  
Tencent Media Lab  
Palo Alto, CA 94306, USA  
zeqiang.work@gmail.com

Abstract—Learning-based visual data compression and analysis have attracted great interest from both academia and industry recently. More training as well as testing datasets, especially good quality video datasets are highly desirable for related research and standardization activities. A UHD video dataset, referred to as Tencent Video Dataset (TVD), is established to serve various purposes such as training neural network-based coding tools and testing machine vision tasks including object detection and segmentation. This dataset contains 86 video sequences with a variety of content coverage. Each video sequence consists of 65 frames at 4K (3840x2160) spatial resolution. In this paper, the details of this dataset, as well as its performance when compressed by VVC and HEVC video codecs, are introduced.

Keywords—4K Dataset, Video Compression, Machine Learning, Video Coding for Machines, Object Detection, Object Segmentation

I. INTRODUCTION

In recent years, new machine learning technologies have brought significant improvements on visual data compression and analysis. In these areas, deep neural networks (DNN) or in general learning-based methods enabled advanced compression efficiency and machine task accuracy through a data driven approach. Efficient learning-based methods are typically designed by going through an extensive training process. Good quality video datasets are therefore highly desirable for related research and standardization activities. JPEG AI, JVET Neural Network based Video Coding (NNVC) and MPEG Video Coding for Machines (VCM) are just a few examples. A UHD video dataset, referred to as the Tencent Video Dataset (TVD) [1], is established in this paper to serve various purposes such as training neural network-based coding tools and testing machine vision tasks including object detection and segmentation.

This dataset contains 86 video sequences with a variety of content coverage. Each video sequence consists of 65 frames at 3840x2160 spatial resolution. This video dataset has been used as a training set in JVET NNVC exploration activities.

For object detection/segmentation explorations, 166 images are sampled from this dataset at 1920x1080 spatial resolution in rgb24 PNG format. Bounding box annotations for these images are provided. These images with annotation have been included in the common test conditions of MPEG VCM as a testing set.

II. VIDEO DATA COLLECTION

All the sequences in the proposed video dataset are captured using Red Helium 8k, Red Monstro 8K and Blackmagic URSA Mini Pro 12K. The sequences are transcoded and then converted to YUV 4:2:0 color format using FFmpeg [2]. Higher resolution format of the same content is possible as the some of the source video clips are captured in a higher resolution (at least 8K).

A variety of scenes with static or moving objects are included in the TVD dataset. The thumbnails of the entire video dataset are provided on the dataset website [1]. In Fig. 1, some sample frames of the sequences in the proposed dataset are presented.

III. EVALUATION ON VIDEO COMPRESSION TOOLS

The rate-distortion behavior of a video sequence is of interest when it is compressed by a typically used video codec. Various video contents contain quite different spatial and temporal characteristics. Also, it is of interest to develop new coding tools for contents that are not very easy to compress. Therefore, training or testing materials that are content rich and challenging to existing video coding solutions can be good for developing new technologies in the field of video compression.

HEVC [3] and VVC [4] are two of the most symbolic video coding standards that are well-known to researchers and engineers. To evaluate the rate-distortion performance of the TVD dataset, video encoding and decoding using VVC reference software version VTM-11.0 [5] and HEVC reference software version HM 16.23 [6] are performed. The overall result statistics as well as some selected sequences are reported.

A. Compression results and comparison

With configurations of AI (All intra)/RA (Random Access)/LDB (Low Delay B)/LDP (Low Delay P), and QP equals to {22, 27, 32, 37, 42}, VTM-11.0 and HM-16.23 are launched for video encoding/decoding on the TVD dataset, using the default encoding parameters in the software packages. In the following tables, the average statistics of the entire dataset are presented. For HM-16.23, the average of PSNR for Y/U/V channel for different QPs is showed in Table I and II. For VTM-11.0, the average of PSNR for Y/U/V channel for different QPs is showed in Table III and IV.
Fig. 1: Some sample frames of 21 example sequences from the TVD dataset
TABLE I.
AVERAGE PSNR FOR YUV CHANNEL FOR DIFFERENT QP WITH HM FOR THE DATASET (AI/RA)

| QP | Bitrate | Y-PSNR | U-PSNR | V-PSNR | Y-PSNR | U-PSNR | V-PSNR |
|----|---------|--------|--------|--------|--------|--------|--------|
| 22 | 707.13  | 47.92  | 47.56  | 48.24  | 10547.61 | 47.71 | 47.48  | 48.17 |
| 27 | 5648.74 | 46.44  | 46.30  | 47.00  | 4058.81  | 46.47 | 46.78  | 47.50 |
| 32 | 3942.50 | 43.85  | 44.79  | 45.63  | 2132.87  | 44.06 | 45.24  | 46.07 |
| 37 | 1821.47 | 41.15  | 43.83  | 44.97  | 1114.81  | 41.41 | 44.84  | 44.80 |
| 42 | 6702.65 | 38.46  | 42.42  | 43.46  | 609.31   | 38.75 | 42.69  | 44.01 |

TABLE II.
AVERAGE PSNR FOR YUV CHANNEL FOR DIFFERENT QP WITH HM FOR THE DATASET (LDB/LDP)

| QP | Bitrate | Y-PSNR | U-PSNR | V-PSNR | Y-PSNR | U-PSNR | V-PSNR |
|----|---------|--------|--------|--------|--------|--------|--------|
| 22 | 977.48  | 49.49  | 48.55  | 49.16  | 8536.60  | 49.15 | 48.65  | 49.36 |
| 27 | 489.92  | 47.62  | 46.70  | 47.62  | 3649.97  | 47.27 | 47.38  | 48.29 |
| 32 | 2698.24 | 44.84  | 45.05  | 46.03  | 1832.88  | 45.00 | 46.01  | 46.96 |
| 37 | 1408.37 | 41.91  | 43.73  | 44.64  | 958.89   | 42.54 | 43.37  | 44.24 |
| 42 | 7299.66 | 39.23  | 42.47  | 43.50  | 503.15   | 39.96 | 43.17  | 44.54 |

TABLE III.
AVERAGE PSNR FOR YUV CHANNEL FOR DIFFERENT QP WITH VTM FOR THE DATASET (AI/RA)

| QP | Bitrate | Y-PSNR | U-PSNR | V-PSNR | Y-PSNR | U-PSNR | V-PSNR |
|----|---------|--------|--------|--------|--------|--------|--------|
| 22 | 977.48  | 49.49  | 48.55  | 49.16  | 8536.60  | 49.15 | 48.65  | 49.36 |
| 27 | 489.92  | 47.62  | 46.70  | 47.62  | 3649.97  | 47.27 | 47.38  | 48.29 |
| 32 | 2698.24 | 44.84  | 45.05  | 46.03  | 1832.88  | 45.00 | 46.01  | 46.96 |
| 37 | 1408.37 | 41.91  | 43.73  | 44.64  | 958.89   | 42.54 | 43.37  | 44.24 |
| 42 | 7299.66 | 39.23  | 42.47  | 43.50  | 503.15   | 39.96 | 43.17  | 44.54 |

TABLE IV.
AVERAGE PSNR FOR YUV CHANNEL FOR DIFFERENT QP WITH VTM FOR THE DATASET (LDB/LDP)

| QP | Bitrate | Y-PSNR | U-PSNR | V-PSNR | Y-PSNR | U-PSNR | V-PSNR |
|----|---------|--------|--------|--------|--------|--------|--------|
| 22 | 881.02  | 48.27  | 47.87  | 48.83  | 9133.67  | 48.21 | 47.84  | 48.59 |
| 27 | 277.54  | 45.98  | 46.49  | 47.37  | 2860.55  | 45.91 | 46.46  | 47.34 |
| 32 | 1307.67 | 43.49  | 43.69  | 46.01  | 1334.70  | 43.46 | 44.07  | 45.07 |
| 37 | 671.52  | 40.84  | 43.55  | 44.50  | 679.89   | 40.91 | 43.52  | 44.47 |
| 42 | 341.03  | 38.31  | 42.32  | 43.39  | 343.80   | 39.28 | 42.29  | 43.26 |

B. Comparison between VTM-11.0 and HM-16.23

The overall BD-Rate [7] performances between VTM-11.0 and HM-16.23 are compared, as shown in Table V. It is demonstrated through the BD-Rate changes that VTM-11.0 can provide a better BD-Rate result over HM-16.23 among all YUV channels for the TVD dataset. Roughly about 33% BD-rate reduction under RA configuration can be achieved using the VTM reference software, when compared with the HM software. The distortion metric in the tests is selected as PSNR.

TABLE V.
BD-RATE GAIN OF VTM-11.0 OVER HM-16.23 ON THE PROPOSED DATASET

| Testing Configuration | BD-rate (VTM-11.0 over HM-16.23) |
|-----------------------|----------------------------------|
| All Intra             | -27.77% -24.65% -25.96% 557% 173% |
| Random Access         | -32.97% -35.95% -37.72% 210% 144% |
| Low Delay B           | -30.50% -31.14% -32.39% 203% 135% |
| Low Delay P           | -34.52% -34.38% -36.98% 199% 137% |

Although more data can be seen, showing the RD curves of the TVD dataset requires too much work and may not be as informative as expected. As showed in Fig. 2 through Fig. 5 in this paper, the RD curves of a set of selected sequences in the dataset are presented to illustrate the performance results comparison between VTM-11.0 and HM-16.23 for video encoding/decoding on the dataset, where configurations of AI (All intra), RA (Random Access), LDB (Low Delay B) and LDP (Low Delay P) are evaluated, respectively. More specifically, four sequences with highest bitrates and four sequences with lowest bitrates are chosen to be presented. For sequences of BoyWithCostume, GirlRunningOnGrass, TreeAndLeaves and BuildingTouristAttraction, they are considered to be among the sequences with highest bitrates in the TVD dataset, when compressed with the same QP value. For these cases, most of the tested points have their PSNR ranges in between 30 and 45dB. On the other hand, sequences of ChefCooking5, FilmMachine, SunriseMountainHuang and GrassLand are among the sequences with lowest bitrates in the dataset, when compressed with the same QP. For these cases, most of the tested points have their PSNR ranges in between 40 and 55dB.

Fig. 2.1: RD curves comparison between VTM-11.0 and HM-16.23 for high bitrate sequences (AI)

Fig. 2.2: RD curves comparison between VTM-11.0 and HM-16.23 for low bitrate sequences (AI)

Fig. 3.1: RD curves comparison between VTM-11.0 and HM-16.23 for high bitrate sequences (RA)

Fig. 3.2: RD curves comparison between VTM-11.0 and HM-16.23 for low bitrate sequences (RA)
From the above figures, the most difficult (high bitrate) sequences operate at as high as 60 Mbps for RA configuration when compressed with QP22; on the other hand, the set of simple (low bitrate) sequences can be compressed at below 3 Mbps. The dramatic differences in bitrate represent the diversity of the contents in the TVD dataset. From Table I–IV, on average, the high-end operating bitrate is around 8–9 Mbps, which are considered to be reasonable for the 4K video contents.

IV. COPYRIGHT INFORMATION

All intellectual property rights of the proposed video clips in the TVD dataset remain with content owners. The following usages of the TVD dataset are allowed and encouraged for the contributed sequences:

- Sequences may be published in technical papers, played at technology research and development events.
- Sequences may be used by standards activities. (e.g., ITU, MPEG, VQEG).
- The following uses are NOT allowed for the contributed sequences:
  - Do not publish snapshots in product brochures.
  - Do not use video for marketing purposes.
  - Redistribution is not permitted.
  - Do not use in television shows, commercials, or movies.

For more detailed terms and conditions, please refer to the TVD dataset downloading website [1].

V. CONCLUSION

This paper presents a new video UHD dataset. Diverse contents are included in this dataset with 4K resolution. It has been evaluated by two different video codecs (VVC and HEVC) with four coding configurations. This dataset fits in various purposes in machine learning related research and standardization activities. Specifically, it has been used in training for NN-based video compression algorithms. With proper annotations, this dataset is further utilized as a test set in video analysis and machine-oriented tasks such as object detection, object segmentation and object tracking.

REFERENCES

[1] Tencent Video Dataset (TVD), available at: https://multimedia.tencent.com/resources/tvd
[2] FFmpeg: A complete, cross-platform solution to record, convert and stream audio and video. https://www.ffmpeg.org/
[3] G. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE Transactions on Circuits and Systems for Video Technology, Volume: 22, Issue: 12, Page(s): 1649 – 1668, Dec. 2012.
[4] Versatile Video Coding. Rec. ITU-T H.266 and ISO/IEC 23090-3, Aug. 2020.
[5] VVC reference software version 11.0: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-11.0
[6] HEVC reference software version 16.23: https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tags/HM-16.23
[7] G. Bjontegaard, “Calculation of Average PSNR Differences Between RD curves”, VCEG-M33, ITU-T SG16 Q.6 document, Austin, TX, USA.