 ISO/IEC JTC 1/SC 29/AG 5 N103

**ISO/IEC JTC 1/SC 29/AG 5**

**MPEG Visual Quality Assessment**

**Convenorship: DE**

**Document type:** General

**Title:** **CVQM - Dataset of compressed video for study of quality metrics (draft 2)**

**Status:** Approved

**Date of document:** 2023-12-08

**Source:** ISO/IEC JTC 1/SC 29/AG 5

**Expected action:** Information

**Action due date:** None

**No. of pages:** 7 (without this cover page)

**Email of Convenor:** wien@lfb.rwth-aachen.de

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag5>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/AG 5**

**MPEG VISUAL QUALITY ASSESSMENT**

**ISO/IEC JTC 1/SC 29/AG 5 N103**

**Hannover, DE – October 2023**

|  |  |
| --- | --- |
| **Title** | CVQM - Dataset of compressed video for study of quality metrics (draft 2) |
| **Source** | AG 5 MPEG visual quality assessment |
| **Status** | Approved |
| **Serial Number** | 23269 |
| **Authors** | Yan Ye (Alibaba), Zhibo Chen (University of Science and Technology of China), Mathias Wien (RWTH Aachen University) |

# Introduction

Exploring accurate and generalized objective video quality metrics is essential for the development of video coding standards, as well as for quality assurance in the multimedia content production and delivery ecosystem. The effectiveness of objective video quality metrics is usually evaluated by comparing them to the ground-truth subjective video quality scores, typically measured as mean opinion scores (MOS). Hence, it is essential to maintain a dataset of original video data and their corresponding coded versions for which ground truth subjective test results are available. There are already some published datasets to study compressed video quality assessment, including Waterloo 4K VQA [1][2], BVI-HD [3], MCL-V [4], CGVDS [5], SCVD [6], and PEA265 [7]. These datasets include MOS scores and objective metrics for video sequences with various characteristics (such as resolution, frame rate, scene representation, etc.) that are coded with different codecs at various coded quality.

However, these datasets are not suitable for AG 5’s work, because some of their copyright terms do not permit them to be used in standard development organizations such as MPEG. Further, AG 5’s Ad hoc Group on learning-based quality metrics is tasked with studying objective quality metrics suitable for future codecs that may rely on learning-based methods, whereas these existing datasets only use traditional video codecs.

To solve this problem, a dataset with diverse content and compression distortion types was proposed in [8]. Following discussions at the 12th meeting of AG 5 in July 2023, it was agreed to establish a dataset of compressed video for study of quality metrics, referred to as the CVQM dataset hereafter. This document provides detailed information about the original and compressed content in the CVQM dataset, including the sources and characteristics of the original content, coding parameters used to obtain the coded video data and objective quality metrics of the coded content.

In version 1 of this output document, only objective quality metrics based on first round of encoding are provided. In future versions, subjective viewing is expected to be conducted, and ground-truth MOS data will be included. Further refinements to encoding parameters are also expected.

In version 2 of this output document, further analyses of the uncompressed video content in CVQM are provided, and more quantization parameters (QPs) are used to generate the compressed video data to cover a wider range of reconstructed qualities and coded bitrates. All bitstreams are re-encoded and cross-checked. Objective metrics are also re-calculated and cross-checked.

# Uncompressed video content in CVQM

A set of 33 original video sequences from Youku, the JVET ftp site and Waterloo 4K datasets are included in the CVQM dataset. These sequences are selected from a total of 108 original video sequences using the methodology described in 2.2. Among these 33 selected CVQM sequences, 2 of them are from Alibaba (including 1 from Alibaba’s Youku), 8 of them are from Waterloo 4K and 23 of them are from JVET. Among the 23 JVET sequences, 6 of them are used in various JVET common test conditions (CTC) and 18 are not used in JVET CTC. Detailed information about these 33 sequences is provided in the accompanying excel (CVQM\_sequeuences.xls). All of these sequences can be obtained from the following site: <https://vqa.lfb.rwth-aachen.de/index.php/apps/files/?dir=/CVQM>.

Only accredited MPEG members can access the ftp site, and login information can be obtained by sending request to AG 5 Convenor Mathias Wien at [wien@lfb.rwth-aachen.de](mailto:wien@lfb.rwth-aachen.de).

## Copyright statements of the original content

### Waterloo 4K content copyright

The copyright statement of Waterloo 4K dataset can be found at [https://ivc.uwaterloo.ca/database/4KVQA.html](https://ivc.uwaterloo.ca/database/4KVQA.htmlf), and is copied and pasted below. It provides reasonable terms of use by standard development organization such as MPEG, and upon our communication with the original authors, they have given explicit permission for AG 5 to use their video data.

*“Permission is hereby granted, without written agreement and without license or royalty fees, to use, copy, modify, and distribute this database (the images, the results and the source files) and its documentation for any purpose, provided that the copyright notice in its entirity appear in all copies of this database, and the original source of this database, Image and Vision Computing Laboratory (IVC, https://ivc.uwaterloo.ca/) at the University of Waterloo (UW, http://www.uwaterloo.ca), is acknowledged in any publication that reports research using this database.”*

It is hereby acknowledged that Waterloo 4K content is provided with the following publication [2]:

* Zhuoran Li, Zhengfang Duanmu, Wentao Liu, Zhou Wang, “AVC, HEVC, VP9, AVS2 or AV1? — A Comparative Study of State-of-the-art Video Encoders on 4K Videos”, 16th International Conference on Image Analysis and Recognition, Waterloo, Ontario, Canada, August 27-29, 2019.

### Youku content copyright

Youku sequence copyright is provided in an earlier JVET contribution JVET-Z0156 [10], copied and pasted below:

*“The proposed sequences are produced by Alibaba Group（licensed by one of its subsidiary Youku） and all intellectual property rights remain with Alibaba Group.*

*The following uses are allowed for the contributed sequences:*

*1. Sequences may be published in technical papers, played at technology research and development events.*

*2. Sequences may be used by Standards committees. (e.g., ITU, MPEG, VQEG).*

*The following uses are NOT allowed for the contributed sequences:*

*1. Do not publish snapshots in product brochures.*

*2. Do not use video for marketing purposes.*

*3. Do not redistribute video with a commercial product.*

*4. Do not use in television shows, commercials, or movies.*

*5. Do not sublicense or transfer the sequences to anyone who has not received this license.”*

### JVET content copyright

Throughout the years, various organizations have provided content for the use of standards development to the joint teams of ITU-T SG16 and ISO/IEC JTC 1/SC 29. These sequences are stored on the JVET ftp site with accompanying copyright statements that permit the use by MPEG.

## Sequence selection methodology

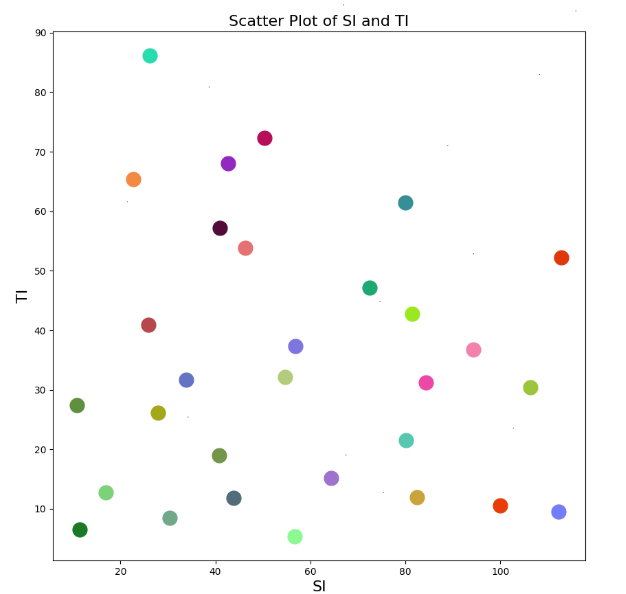
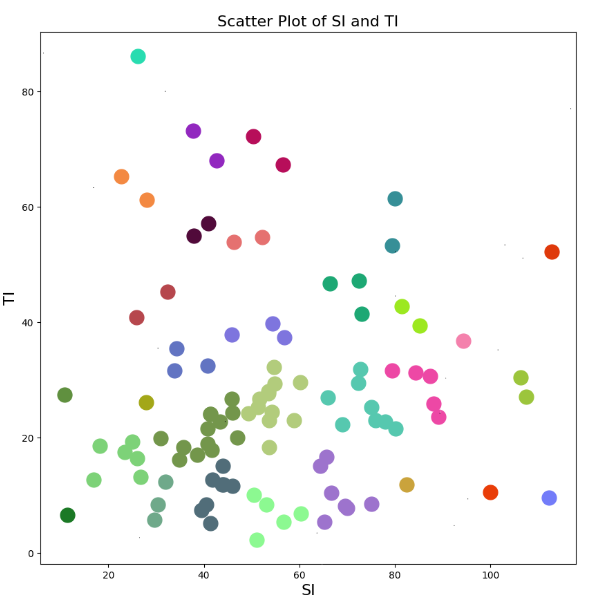
The sequence selection process started from a total of 108 candidate sequences from Youku, JVET ftp site and Waterloo 4K datasets. All these sequences are in YUV420 format, and have a duration of at least 10 seconds. If a sequence is longer than 10 seconds, then a 10-second segment is cropped from it to become the candidate sequence.

All candidate sequences are annotated with scene categories, encompassing a wide range of commonly encountered video compression scenes. There are 15 scene categories in total, including animal, building, cartoon, crowd, game, indoor activities, marketplace, natural landscape, night scene, object, outdoor scene, outdoor (other), people (one or more), screen content, and sports.

The degree of distortion in the compressed video is related to the spatial-temporal complexity of the content, which can be measured by two commonly used attributes, spatial information (SI) and temporal information (TI). Therefore, sequences in the CVQM dataset should have wide coverage of SI and TI values.

To select appropriate sequences from these 108 candidate sequences, the following steps are followed:

1. The SI and TI values of the 108 candidate sequences are calculated using the VQEG SITI-Tools (available at <https://github.com/VQEG/siti-tools>). Figure 1 (a) shows the SI and TI distribution of all 108 candidate sequences;
2. Based on the SI/TI scatter plot, a "cluster then select" paradigm is applied to down select a smaller set of representative sequences from the 108 candidates. The Agglomerative Clustering method [11] is used for clustering. The sequences belonging to the same cluster have similar SI/TI values, therefore one is selected and other redundant samples are removed. This is performed in an iterative manner, and finally a smaller set of intermediate sequences are selected. As shown in Figure 1 (b), these intermediate sequences cover a wide SI/TI range and the distribution exhibits good uniformity;



1. （b）

Figure 1. SI/TI distribution of (a) all 108 candidate sequences, and (b) intermediate selected sequences

1. Further minor adjustments are performed to ensure that all 15 scene categories are represented by the dataset. Figure 2 depicts the final selected sequences in the CVQM dataset, their respective category labels, and the range of SI/TI values they cover. There are 33 video sequences in total. Night scene and outdoor scene contain multiple samples. The samples from night scene are concentrated in the low SI/TI region, which is consistent with the observation that these night scene videos contain lower color richness and lower temporal fluctuations. For outdoor scene, the samples cover a wider SI/TI range, which reflects the richness of the space-temporal complexity in the videos in this category.

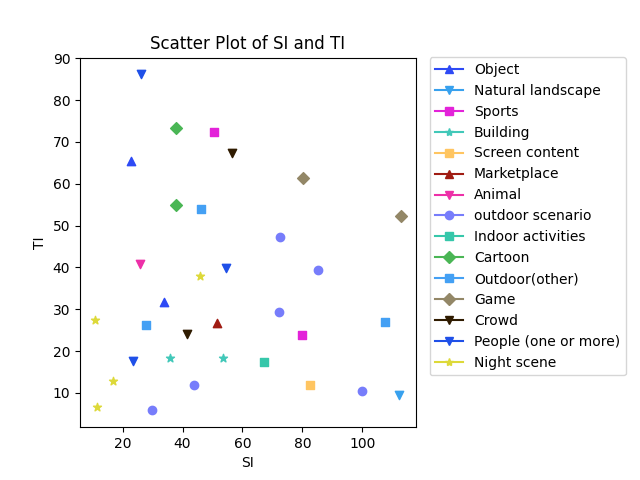


Figure 2. Annotation of the 33 original videos in the CVQM dataset, as well as their SI/TI distribution

## Characteristics of the uncompressed video content

Besides SI and TI analysis shown in Figure 2, Video Complexity Analyzer (VCA) [12] is applied to the CVQM content. VCA is an open-source tool designed to assess spatial and temporal complexity by relying on the Discrete Cosine Transform (DCT) energy. Among the seven block-wise DCT-based energy features, visualization of VCA’s spatial complexity “E” and temporal complexity “H” is provided in Figure 3. It can be seen that the CVQM content has sufficient coverage in the VCA space.

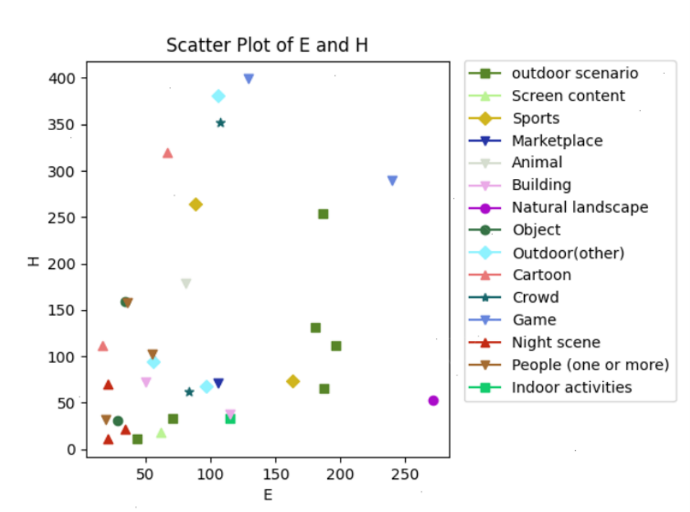
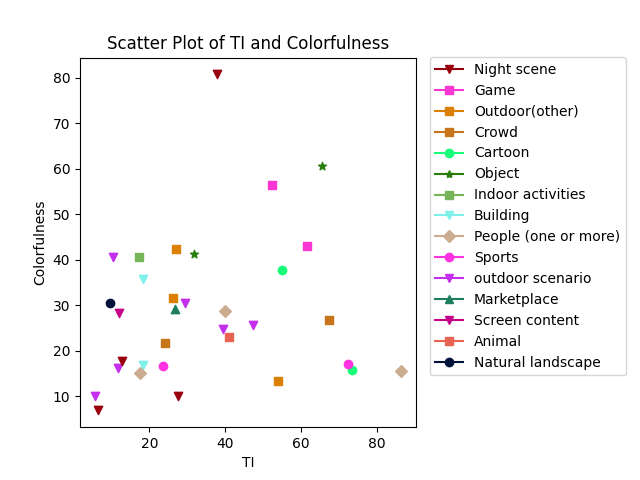
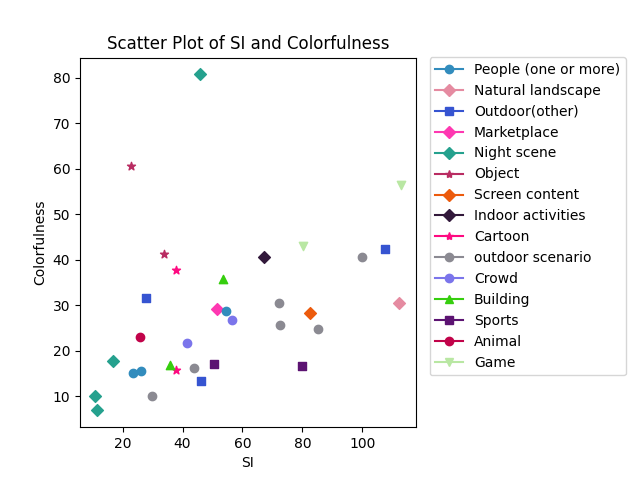


Figure 3. Distribution of VCA’s spatial and temporal complexity (“E” and “H”) of the CVQM sequences

Colorfulness (CF) [12][13] is commonly used to measure the variety and intensity of colour information in the images. The CF vs. SI and CF vs. TI plots for the CVQM content are provided in Figure 4 (a) and Figure 4 (b), respectively.



(a) (b)

Figure 4. CF vs SI distribution (a) and CF vs TI distribution (b) of the CVQM sequences

The presence of scene cut(s) in a video could affect the coded bitrates and perceived quality of the reconstructed video due to temporal motion discontinuity. Scene cuts could be the result of a sudden change between two scenes (e.g. due to sudden change in camera angle) or a more gradual change between two scenes (e.g. due to the use of fade-in, fade-out, or cross-fade techniques). Whether scene cuts are present, the number of scene cuts if present, and the type of scene cuts if present, are recorded in the accompanying spreadsheet AG05N0103\_CVQM\_sequences\_v2.xls.

# Compressed video data in CVQM

For study of learning-based quality metrics, it is desirable to include coded video that has been compressed using learning-based video codecs in addition to those compressed using traditional codecs. In Draft 1 of this output document, it was agreed to include only traditional codecs. Further study on potential inclusion of learning-based video codecs is needed and planned. Two well-known ITU-T and MPEG codecs, namely H.265/HEVC and H.266/VVC, are selected to produce the coded video data in CVQM. Two implementations per codec are used, one from reference software and the other from open source. For H.265/HEVC, [HM-18.0](https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tags/HM-18.0) and [x265-3.5+103-8f18e3a](https://bitbucket.org/multicoreware/x265_git/commits/8f18e3ad32684eee95e885e718655f93951128c3) are used. For H.266/VVC, [VTM-20.2](https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-20.2) and [VVenC-1.9.0](https://github.com/fraunhoferhhi/vvenc/releases/tag/v1.9.0) are used.

All coded bitstreams can be obtained from the following site: <https://vqa.lfb.rwth-aachen.de/index.php/apps/files/?dir=/CVQM>.

The HM and VTM are coded with random access configuration. For x265 and VVenC, slower preset and medium preset with QPA enabled are used, respectively. And for all of these encoders, QP are set as {12, 17, 22, 27, 32, 37, 42, 47}. Furthermore, only the first 10 seconds for each sequence have been used for encoding.

The same login information as those for the uncompressed data is needed to access and download these bitstreams.

For objective quality metrics, PSNR, SSIM, MS-SSIM and VMAF are provided along with the coded video data. PSNR is calculated according to the technical paper “Working practices using objective metrics for evaluation of video coding efficiency experiments” as specified in [HSTP-VID-WPOM](https://www.itu.int/dms_pub/itu-t/opb/tut/T-TUT-ASC-2020-HSTP1-PDF-E.pdf). PSNR, SSIM and MS-SSIM metrics are calculated using [HDRTools v0.24](https://gitlab.com/standards/HDRTools/-/commit/b6fde1487851ab417b5a317e03fee6d522d45f45), and VMAF is calculated using [VMAF v2.3.1](https://github.com/Netflix/vmaf/releases/tag/v2.3.1).

All metrics for the coded data are reported in the accompanying excel (AQ05N00103\_CVQM\_coded\_data\_v2.xls).

In v1 of this document, the encoding uses 4 QPs per sequence per encoder. In v2 of this document, to cover a wider range of reconstructed qualities and coded bitrates, the encoding is expanded to using 8 QPs for the HEVC and VVC encoders. All bitstreams are re-encoded and cross-checked. Objective metrics are also re-calculated and cross-checked.

# Other information provided on the CVQM site

The following information can be found on the CVQM site <https://vqa.lfb.rwth-aachen.de/index.php/apps/files/?dir=/CVQM>:

* Readme file with relevant administrative information
* AQ05N00103\_CVQM\_sequences\_v2.xls containing information about the original sequences
* AQ05N00103\_CVQM\_coded\_data\_v2.xls containing information about the coded bitstream (codec used, QP, bit rates, objective quality metrics, etc.)

# References

1. Z. Li, Z. Duanmu, W. Liu, Z. Wang, “Waterloo 4K Video Quality Assessment Database: A Comparative Study of Modern Video Encoders,” https://ece. uwaterloo. ca/~ z777li/4kvqa/.
2. Z. Li, Z. Duanmu, W. Liu, Z. Wang, “AVC, HEVC, VP9, AVS2 or AV1? — A Comparative Study of State-of-the-art Video Encoders on 4K Videos”, 16th International Conference on Image Analysis and Recognition, Waterloo, Ontario, Canada, August 27-29, 2019.
3. F. Zhang, F. M. Moss, R. Baddeley, and D. R. Bull, “BVI-HD: A video quality database for HEVC compressed and texture synthesized content,” IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2620-2630, 2018.
4. J. Y. Lin, R. Song, C.-H. Wu, T. J. Liu, H. Wang and C.-C. J. Kuo, “MCL-V: A streaming video quality assessment database,” Journal of Visual Communication and Image Representation, Vol. 30, pp. 1-9, July 2015.
5. S. Zadtootaghaj, S. Schmidt, S. Shafiee Sabet, S. Moeller, and C. Griwodz, “Quality Estimation Models for Gaming Video Streaming Services Using Perceptual Video Quality Dimensions,” Proceedings of the 11th International Conference on Multimedia Systems. ACM. 2020.
6. S. Cheng, H. Zeng, J. Chen, J. Hou, J. Zhu, K.K. Ma, “Screen content video quality assessment: Subjective and objective study,” IEEE Transactions on Image Processing, Vol. 29, pp. 8636-8651, 2020.
7. L. Lin, S. Yu, L. Zhou, W. Chen, T. Zhao, Z. Wang, “PEA265: Perceptual assessment of video compression artifacts,” IEEE Transactions on Circuits and Systems for Video Technology, 30(11), pp. 3898-3910, 2020.
8. Y. Lu, F. Guan, Y. Qi, Y. Ren, X. Li, Z. Yu, Z. Chen, R.-L. Liao, Y. Ye, “Video dataset for study of learning-based quality metrics,” ISO/IEC JTC 1/SC 29/AG 5, document no. m64283, July 2023.
9. S. Wang, M. Wien and Y. Ye, "Preliminary databases for assessment of quality metrics for 2D video using neural-network (NN) based methodologies," ISO/IEC JTC 1/SC 29/AG 5, document no. m57121, July 2021.
10. J. Chen, Y. Ye, R. Li, W. Jiang, “AHG4: New test sequences for JVET exploration,” the joint video experts team (JVET) of ITU-T SG16 and ISO/IEC JTC 1/SC 29, document no. JVET-Z0156, April 2022.
11. D. Müllner, “Modern hierarchical, agglomerative clustering algorithms,” arXiv preprint arXiv:1109.2378, 2011.
12. Winkler S. Analysis of public image and video databases for quality assessment[J]. IEEE Journal of Selected Topics in Signal Processing, 2012, 6(6): 616-625.
13. D. Hasler and S. E. Suesstrunk, “Measuring colorfulness in natural images,” in Proc. SPIE Human Vision and Electronic Imaging, vol. 5007, Jun. 2003, pp. 87–95.

# Annex A: List of CVQM sequences

1. The sequence “BGStatic2Moving\_OnePeople” is from JVET-T0060 and the sequence “PostalService” is from JVET-Z0156.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Video Name** | **source** | **width** | **height** | **frame rate** | **Bit depth** | **# of frames** | **scene cut** | **MD5** |
| Waterloo\_01 | Waterloo | 3840 | 2160 | 24 | 8 | 240 | No | 015de9df9398279ec2988e67ef4d7ad8 |
| Waterloo\_02 | Waterloo | 3840 | 2160 | 25 | 8 | 250 | No | 3dc18fd4da4584cb9db067b0aafa5559 |
| Waterloo\_04 | Waterloo | 3840 | 2160 | 24 | 8 | 240 | No | ec40855a65b104e0529c9edf937b5eb6 |
| Waterloo\_05 | Waterloo | 3840 | 2160 | 30 | 8 | 300 | No | 017b0d2cc3c5cd03b3c90c7a72074271 |
| Waterloo\_06 | Waterloo | 3840 | 2160 | 30 | 8 | 300 | No | d798d32ff68ff24045cd1e5d5578b410 |
| Waterloo\_11 | Waterloo | 3840 | 2160 | 30 | 8 | 300 | Yes | 5b09e998e8a1059af962d7db23e4ba94 |
| Waterloo\_12 | Waterloo | 3840 | 2160 | 30 | 8 | 300 | Yes | c16169fe2f1593c56007e48dc1bae6cd |
| Waterloo\_18 | Waterloo | 3840 | 2160 | 25 | 8 | 250 | No | 99a20338575a84b9b84694fb5dad2052 |
| 3DGameKit4K | JVET | 3840 | 2160 | 60 | 10 | 601 | No | f92dde781b7213821e89b03ef77eaa83 |
| BasketballDrive | JVET | 1920 | 1080 | 50 | 8 | 501 | No | d38951ad478b34cf988d55f9f1bf60ee |
| BQTerrace | JVET | 1920 | 1080 | 60 | 8 | 601 | No | efde9ce4197dd0b3e777ad32b24959cc |
| Cactus | JVET | 1920 | 1080 | 50 | 8 | 500 | No | 3fddb71486f209f1eb8020a0880ddf82 |
| Campfire | JVET | 3840 | 2160 | 30 | 10 | 300 | No | 63d3d9f9e4e8b5c344e89840e84e6428 |
| DroneTakeOff | JVET | 3840 | 2160 | 30 | 10 | 300 | No | b787e95fb3d3e066f4717ccf8875f972 |
| Fallout4 | JVET | 1920 | 1080 | 60 | 8 | 3602 | Yes | 09980ed191d285449d591fb0cf3dbe43 |
| FoodMarket4 | JVET | 3840 | 2160 | 60 | 10 | 720 | No | a378b34190f54f688d048a9a8b46a8ac |
| Fountains | JVET | 3840 | 2160 | 30 | 8 | 300 | No | 662ce078ab5bcfa068b6b5f45f5cd930 |
| NeptuneFountain3 | JVET | 3840 | 2160 | 60 | 10 | 600 | No | 88fd87ea57df4a36200946025e8618aa |
| Netflix\_Dancers | JVET | 4096 | 2160 | 60 | 10 | 1199 | No | e440efb517c01d27ecf072c29e31197b |
| Netflix\_DinnerScene | JVET | 4096 | 2160 | 60 | 10 | 1199 | No | 4cf8bb5eb4610ab20266332a638c3c55 |
| Netflix\_Meridian2 | JVET | 1920 | 1080 | 60 | 10 | 600 | Yes | 473757104c47837efe3fc3b85aef642a |
| Netflix\_RollerCoaster | JVET | 4096 | 2160 | 60 | 10 | 1199 | No | 8e99b7f486025ce8fccda4f3e818a186 |
| Netflix\_SquareAndTimelapse | JVET | 4096 | 2160 | 60 | 10 | 600 | No | a7ffe67fde0a41b313c7cc9317195d44 |
| Netflix\_ToddlerFountain | JVET | 4096 | 2160 | 60 | 10 | 1199 | Yes | 31eb61104464922d91f42e5d5b49097e |
| NightRoad | JVET | 3840 | 2160 | 60 | 10 | 600 | No | e13f46ff51607fd5eda429cd6f9b3d2d |
| OberbaumSpree\_ | JVET | 3840 | 2160 | 60 | 10 | 600 | No | 6975b81c9e63c92b3bf4223796102da1 |
| ResidentialBuilding | JVET | 3840 | 2160 | 30 | 8 | 300 | No | e1b3a496b20a88fa88f7cd2e0261c68a |
| RitualDance | JVET | 1920 | 1080 | 60 | 10 | 600 | Yes | a3cb399a7b92eb9c5ee0db340abc43e4 |
| Stem | JVET | 1920 | 1080 | 25 | 8 | 250 | No | 9e17f565a8c1beb46a5bc02e680c64ff |
| TiergartenParkway | JVET | 3840 | 2160 | 60 | 10 | 600 | No | cc17b64e1fb93879c0873444ca13e290 |
| Wood | JVET | 3840 | 2160 | 30 | 8 | 300 | No | d7c6e0954971b4d08a4de724f06f5e44 |
| BGStatic2Moving\_OnePeople | Alibaba1 | 1920 | 1080 | 30 | 8 | 450 | No | fe3b72f9f4fd3474a93b66e4db10c4be |
| PostalService | Alibaba1 | 3840 | 2160 | 25 | 8 | 375 | No | 203db979f31f431c6b9f7afe0918892d |