 **ISO/IEC JTC 1/SC 29/ WG 11 W 19507**

**ISO/IEC JTC 1/SC 29/WG 11**

**Coding of moving pictures and audio**

**Convenorship: Japan (JISC)**

|  |  |
| --- | --- |
| **Document type:** | Approved WG 11 document |
| **Title:** | **Draft Evaluation Framework for Video Coding for Machines** |
| **Status:** | Approved |
| **Date of document:** | 2020-07-03 |
| **Source:** | Convenor, ISO/IEC JTC 1/SC 29/WG 11 |
| **No. of Pages:** |  |
| **Email of acting convenor** | ostermann@tnt.uni-hannover.de |
| **Committee URL:** | <http://isotc.iso.org/livelink/livelink/open/jtc1sc29> |
|  |  |

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11/w****19507**

**June-July 2020, Online Meeting**

**Title Draft Evaluation Framework for Video Coding for Machines**

**Source MPEG Requirements**

**Status:**

**Table of Contents**

[1 Introduction 4](#_Toc44667062)

[2 Test Conditions 4](#_Toc44667063)

[3 Anchor Definition and Requirements 5](#_Toc44667064)

[3.1 NN Tasks 6](#_Toc44667065)

[3.2 Network architectures 6](#_Toc44667066)

[3.3 Datasets . 6](#_Toc44667067)

[4 Evaluation Methods and Procedures 9](#_Toc44667068)

[4.1 Proposed Processing Pipelines 10](#_Toc44667069)

[4.2 Pre/Post-Processing Data (image) Conversion 10](#_Toc44667070)

[4.2.1 Processing Pipeline for Downscaled Resolution 11](#_Toc44667071)

[4.3 Anchor Reference Compression / Decompression 12](#_Toc44667072)

[4.4 Performance Evaluation 13](#_Toc44667073)

[5 Evaluation Approach for Machine and Human Consumption 14](#_Toc44667074)

[5.1 Complexity Measurement 16](#_Toc44667075)

[6 References 17](#_Toc44667076)

[7 Appendix A: Anchor Generation Results 18](#_Toc44667077)

[7.1 Object Detection 18](#_Toc44667078)

[7.2 Object Segmentation 25](#_Toc44667079)

[7.3 Video Segmentation 37](#_Toc44667080)

[8 Appendix B: Anchor Runtime Measurement 41](#_Toc44667081)

# Introduction

The MPEG activity on Video Coding for Machines (VCM) aims to standardize a bitstream format generated by compressing either a video stream or previously extracted features. The bitstream should enable multiple machine vision tasks. VCM shall be able to

* Efficiently compress the bitstream; the size of the compressed features shall be less than the encoded video stream using state-of-the-art video compression technologies like VVC.
* Use the bitstream to support single or multiple tasks. Decompressed bitstream should be general enough to be usable for different scenarios, for example object detection and segmentation.
* Support varying performance for multiple tasks as measured by the appropriate metrics. This performance level may depend on the application.
* Allow the reconstruction of the compressed bitstream for human consumption. This can be achieved with an additional bitstream.

MPEG VCM has identified a set of relevant use cases and related requirements [1], focusing on the machine-to-machine communication in intelligent transportation and intelligent industry; and the hybrid machine and human consumption for surveillance and smart city use cases. This document contains information on how to provide evidence for these use cases. [2]

* Datasets: which datasets should be used for which sub-tasks, where these datasets can be obtained, how the datasets are split into training and validation data
* Metrics: which metric shall be used for which sub-tasks, how these metrics are calculated, what to compare performance results against

# Test Conditions

Decoded video/feature shall be tested for one or more key tasks for a specific use case and compare the performance results to current anchors. Retraining the shared backbone is permitted using joint training or other approaches in the case of two or more key tasks. Modifications and training of the task-specific networks are allowed but need to be reported in detail. In some cases, the encoder may know the task-specific neural networks at the decoder side. In this document, framework refers to the used datasets and software packages.

# Anchor Definition and Requirements

VVC/H.266 codec with software version VTM-8.2 is used as the reference for the performance evaluation of MPEG-VCM encoder. Table 1 shows the tasks along with their metrics, datasets, benchmarks, and training/testing description.

Table *Training and test conditions, key metrics, datasets, benchmarks for various tasks*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Metrics** | **Datasets** | **Benchmarks** | **Training/Testing** |
| Object Detection | [mAP](https://mc.ai/the-confusing-metrics-of-ap-and-map-for-object-detection/)  vs  BPP/Rate | COCO [compressed]  (image) | <http://cocodataset.org/#detection-leaderboard> | For COCO, use 2014/2017 Val set for evaluation and 2014/2017 Train in the case of retraining. |
| [CityScapes](https://www.cityscapes-dataset.com/) [uncompressed]  (image)  CityPersons  [uncompressed]  (image) | <https://www.cityscapes-dataset.com/benchmarks/> | For CityScapes, use defined train and validation sets |
| [ImageNet](http://www.image-net.org/) [compressed]  (image) | <https://kobiso.github.io/Computer-Vision-Leaderboard/imagenet.html> | For Imagenet, use the training and validation data as published from ILSVRC 2014. |
| Object Segmentation | [mAP](https://mc.ai/the-confusing-metrics-of-ap-and-map-for-object-detection/)  vs  BPP/Rate | COCO  [compressed]  (image) | <http://cocodataset.org/#detection-leaderboard> | (see above) |
| [CityScapes](https://www.cityscapes-dataset.com/) [uncompressed]  (image) |  | (see above) |
| [KITTI](http://www.cvlibs.net/datasets/kitti/index.php)  (image) | <http://www.cvlibs.net/datasets/kitti/eval_object.php> | We recommend using the predefined splits. |
| [mAP](https://mc.ai/the-confusing-metrics-of-ap-and-map-for-object-detection/)  vs  Rate | DAVIS 2016 / 2017  (video) | <https://davischallenge.org/> | We recommend using the semi-supervised mode for higher accuracy. |
| Object Tracking | MOTA  vs  Rate | MOT20  [compressed]  (video) | <https://arxiv.org/pdf/1906.04567.pdf> | Dataset split is available from the Tracking Challenge, available on their website. |

## NN Tasks

At the current stage of this document, the key NN tasks for the MPEG-VCM performance evaluation are:

* Object Detection (still image)
* Object Segmentation (still image/video)
* Object Tracking (video)

## Network architectures

* Object Detection: Faster R-CNN X101-FPN (part of Facebook AI Research’s Detectron)

* Object Segmentation: R50-FPN Cityscapes (part of Facebook AI Research’s Detectron), PreMOVOS

* Object Tracking: JDE-1088x608
* Optional requirement
  + As an additional performance data, other network architectures are allowed to be used per tasks. However, it is up to the proponents to provide comparable performance data with the anchors specified in this document

## Datasets .

The datasets to be used for anchor generation evaluation framework and VCM standardization process shall be of high quality, available and downloadable, sufficiently modern, have permissive licenses, and support adequate pre-trained models.

* + Datasets must be of high quality such that a VVC encoder can create several coded versions of the sequence with noticeable degradation in quality. It is recommended to use uncompressed datasets where possible.
  + In some datasets, most images are compressed with quality factors around 96 and 90, while some images were even compressed with quality factor around 80. It is recommended to build up a subset of the dataset by removing the lower quality images.
  + Datasets shall be capable of generating anchors per requirements
  + Datasets must be available and downloadable
  + Datasets licensing terms shall allow for using them in standardization and should allow for commercial use
  + Datasets shall be sufficiently modern
  + Datasets that support pre-trained models

Proponents are invited to look into the datasets and raise a flag in case of encountering issues such as copyrights, etc. (Note: It is intended to have these datasets to be downloadable from the MPEG site). Table 2 shows the recommended datasets for VCM anchor generation.

Table *. The following datasets are recommended to be used for anchor generation*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Image /**  **Video** | **Compressed /**  **Uncompressed** | **Benchmarks** | **Comments** |
| COCO | Image | Compressed | <http://cocodataset.org/#detection-leaderboard> | For COCO, use 2017 Val set for evaluation and 2017 Train in the case of retraining  80 object classes  Will check the licensing issue |
| Cityscapes | Image | Uncompressed | <https://www.cityscapes-dataset.com/benchmarks/> | For Cityscapes, use defined train and validation sets  Will check the licensing issue |
| Citypersons | Image | Uncompressed | <https://www.cityscapes-dataset.com/benchmarks/> | For Citypersons, use defined train and validation sets  Will check the licensing issue |
| [KITTI](http://www.cvlibs.net/datasets/kitti/index.php) | Image | Uncompressed | <http://www.cvlibs.net/datasets/kitti/eval_object.php> | We recommend using the predefined splits  Will check the licensing issue |
| DAVIS 2016 / 2017 | Video | Compressed | <https://davischallenge.org/> | We recommend using the semi-supervised mode for higher accuracy.  Will check the licensing issue  Creative Commons Attributions 4.0 License (non-commercial use) |
| MOT20 | Video | Compressed | <https://arxiv.org/pdf/1906.04567.pdf> | Dataset split is available from the Tracking Challenge, available on their website.  Cannot be used for commercial use  (can be used only for academia) |
| UCF101 | Video | Compressed | <https://www.crcv.ucf.edu/data/UCF101.php> | Action recognition  13320 videos from 101 action categories  Will check the licensing issue |
| HiEve | Video | 14 Uncompressed | <http://humaninevents.org/data.html?title=1> | Human in Event  19 training videos, 13 test videos and a total of 32 videos  No issues with licensing – can provide license/permission to MPEG-VCM  Annotation of bounding box, ID, key points and behavior, dense pose. Can be used for multi-target tracking  Proponent agreed to provide permission (licensing free) for MPEG-VCM activities at MPEG131 (by 2020-07-03 - Received) |
| SFU-HW-Objects-v1 | Video | Uncompressed |  | Labeled video data  Object labeled dataset on raw video sequences  Already being used for MPEG (HEVC) – appears to be fine for standardization activity  Need to double-check to see whether can be used for MPEG-VCM  Can be used for compression and object detection simultaneously  Need to be investigated to see whether it large enough to be used for VCM activities  The dataset is provided under the Creative Commons license BY 4.0 (CC BY 4.0)  Proponent agreed to provide permission (licensing free for labels) for MPEG-VCM activities at MPEG131 (by 2020-07-03) |
| Open Image | Image | Compressed |  | 9M images, 600 classes, Modern.  Listed as using CC BY 2.0 and CC BY 4.0 Licensing terms |
| [FLIR Thermal Dataset](https://www.flir.com/oem/adas/adas-dataset-form/) | Image | Compressed | <https://www.flir.com/oem/adas/adas-dataset-form/> | Ability to sense thermal infrared radiation or heat  Will check the licensing issue |

# Evaluation Methods and Procedures

The evaluation procedure and metrics are described in section 2 above. The metrics consist of two parts, one relating to feature extraction and one relating to compression of processed or unprocessed video. The metrics and anchors for feature extraction will be considered later.

The majority of these datasets have publicly defined training and validation sets. In the case this is not available, we will release a training and testing split for comparison. This list is not exhaustive, and proponents are free to use their own datasets for each of the key tasks.

The input images and labels for training and testing are directly taken from the dataset for specific use cases as listed above. This leads into a general feature extractor such as a convolutional neural network, which converts the images or video into a stream of unprocessed or processed video. The resulting features are then fed into different machines, whose results are calculated with respect to the appropriate metric. Proponents are asked to report this result along with the current state of the art on the chosen group of tasks, which will be released by MPEG-VCM. A comparison will be made regarding the performance across the different tasks in the group measured by the relevant metric.

Regarding the compression of processed or unprocessed video, proponents are asked to test the compression ratio on the processed or unprocessed video. This compression ratio should be given as a comparison to the released compression ratio of VVC on the unprocessed video. For human consumption use cases, proponents shall report BD-rate. BD-rate should be calculated in the way as other standardization groups, e.g. JVET [3].

* Usecase specific performance metrics, with the key tasks and metrics as defined in Table 1. Proponents shall perform the evaluation themselves, with the experimental conditions described in [1].
* Compression efficiency, runtime complexity, and memory consumption of compression / decompression (measurement is independent of the use case). Proponents shall perform the evaluation themselves based upon a provided unprocessed or processed video. In the case of processed video, the output may come from a common neural network or general feature extraction methods regarding the specific key tasks. As an example, these common neural networks backbones may be VGG, ResNet, Inception and the specific frameworks depend on the key tasks. For detection and segmentation, an example may be Mask R-CNN or YOLO.

## Proposed Processing Pipelines

The pipeline architecture for anchor generations is shown in Figure 1.a. Possible pipeline architectures for technology proposal are shown in Figure 1.b and 1.c.

|  |
| --- |
|  |
| 1. Pipeline 1 |
|  |
| 1. Pipeline 2 |
|  |
| 1. Pipeline 3 |

Figure . Proposed processing pipeline [4]

## Pre/Post-Processing Data (image) Conversion

For input data processing, it is suggested to use FFmpeg release 4.2.2. FFmpeg can be used for data format conversion, up/down-sampling and resizing (cropping/padding/scaling) the image.

* + FFmpeg 4.2.2
    - Resolution: Scaling / resolution (100%, 75%, 50%, 25%):
* Scale PNG image to new resolution:

ffmpeg -i input.png -vf “scale=NEW\_WDT:NEW\_HGT“ output.png

for 100%: -vf “pad=ceil(iw/2)\*2:ceil(ih/2)\*2”

for 75%: -vf "scale=ceil(iw\*3/8)\*2:ceil(ih\*3/8)\*2"

for 50%:   -vf "scale=ceil(iw\*/4)\*2:ceil(ih\*/4)\*2"

for 25%: -vf "scale=ceil(iw\*/8)\*2:ceil(ih\*/8)\*2"

* + - Format conversion: Convert PNG 🡨🡪 YUV:

ffmpeg -i input.png -f rawvideo -pix\_fmt yuv420p *-dst\_range 1* output.yuv

ffmpeg -f rawvideo -pix\_fmt yuv420p10le -s WDTxHGT -*src\_range 1 -i* input.yuv -frames 1 -pix\_fmt rgb24 output.png

### Processing Pipeline for Downscaled Resolution

There are two possible processing pipelines [5] for the downscaled ground-truth image as are shown in Figure 1. For anchor generations, an alternative processing pipeline shown in Figure 2.b is recommended.

|  |
| --- |
|  |
| 1. Original Pipeline |
|  |
| 1. Alternative pipeline – Upscaled decoded image |

Figure . Possible pipelines for downscaled resolutions

## Anchor Reference Compression / Decompression

VTM8.2 encoder is used to compress generated YUV files under the default “All Intra” configuration and ConformaceWindowMode set to 1. ConformaceWindowMode is equal to 1 represent padding width and height of processing image automatic to multiple of minimal CU size.

A total of six evenly-spaced QPs are recommended, respectively (17 optional), 22, 27, 32, 37, 42, and 47.

* VVC: Reference software VTM-8.2
* JVET Common Test Conditions (CTC-420) with Random Access (RA) condition for videos
* JVET Common Test Conditions (CTC-420) with All Intra AI condition for images
* encoder -c cfg/encoder\_intra\_vtm.cfg -i input.yuv -o reconstruct.yuv -b compress.vvc -q QP --ConformanceWindowMode=1 -wdt WDT -hgt HGT -f 1 -fr 1 --InternalBitDepth=10
* decoder -b compress.vvc -o decode.yuv

## Performance Evaluation

To evaluate the accuracy and compression efficiency of VCM (object detection, object segmentation, and object tracking), mAP/MOTA vs BPP/bitrate curves are to be generated with the following specifications:

* Specify max. endpoint (uncompressed) performance (mAP/MOTA)
* Specify min. endpoint (threshold) performance (mAP/MOTA)
* Checkpoints are uniformly distributed between max and min endpoints (including the endpoints)
* Generate performance (mAP/MOTA v.s. BPP) curves for each task based on the 4x resolutions and QP within the range of acceptable performance (min/max endpoints)
* Performance curve is the Pareto front created from the 4 curves
* Accuracy measurement - mAP
  + mAP, AP@[0.5:0.95]
* Compression efficiency measurement – BPP/bitrate/MOTA
  + BPP calculation: BPP is calculated with respect to the original image resolution (not the downscales image)

# Evaluation Approach for Machine and Human Consumption

The evaluation process of video test data set for machine consumption and human consumption using VVC as an anchor are shown in Figure 3 and Figure 4, respectively.

|  |
| --- |
|  |

Figure . Evaluation approach of video test data using VVC anchor - machine consumption

|  |
| --- |
|  |

Figure . Evaluation approach of video test data using VVC anchor - human consumption

The feature data type and format information are beneficial for compression experts to know the properties and limitations of the feature data types and formats to increase the quality of their proposals. The list of feature data types and formats is recommended to be as exhaustive as possible and to include all relevant information such as allowed values and data ranges.

Table 3 shows an overview of different data types for different tasks required by the use cases described in [1].

Table . *Feature data types and their description for various tasks*

|  |  |  |
| --- | --- | --- |
| **Task** | **Type of data** | **Description** |
| Object detection | List of bounding boxes | Maximum number of bounding boxes: TBD  Each bounding box has four attributes:   * pos\_x (integer): offset from left picture edge: 0 to MAX\_PIC\_WIDTH * pos\_y (integer): offset from top picture edge: 0 to MAX\_PIC\_HEIGHT * size\_x (integer): width from left edge of bounding box: 1 to MAX\_PIC\_WIDTH * size\_y (integer): height from top edge of bounding box: 1 to MAX\_PIC\_HEIGHT |
| Object segmentation | Matrix | Matrix size: INPUT\_WIDTH x INPUT\_HEIGHT  All elements of the matrix are either a single integer value or a list of three integer values (for different color formats). The range of the values depends on the chosen bit depth. |
| Object tracking | List of bounding boxes | Maximum number of bounding boxes: TBD  Each bounding box has five attributes:   * pos\_x (integer): offset from left picture edge: 0 to MAX\_PIC\_WIDTH * pos\_y (integer): offset from top picture edge: 0 to MAX\_PIC\_HEIGHT * size\_x (integer): width from left edge of bounding box: 1 to MAX\_PIC\_WIDTH * size\_y (integer): height from top edge of bounding box: 1 to MAX\_PIC\_HEIGHT   box\_id (integer): identifier for each box to allow tracking through multiple frames |

## Complexity Measurement

Encoder/decoder runtime has served as a reasonable proxy for computational complexity in JVET / VVC standardization projects. It is recommended the runtime metric [6] to be reported for anchors generation in Machine vision and Human vision tasks.

The complexity of the codec shall be independent of pre-processing methods and NN tasks used. For the recommended method to measure the codec complexity see Appendix B.

# References

|  |  |
| --- | --- |
| [1] | w19506, "Use cases and draft requirements for Video Coding for Machines," Online, June 2020. |
| [2] | T.-H. Li, Y.-C. Nien and D.-R. L. (Foxconn), "m54374 [VCM] Anchor results for object segmentation on COCO dataset," June 2020. |
| [3] | K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström and G. J. Sullivan, "Summary information on BD-rate experiment evaluation practices. JVET-Q2016 (w19168)," January 2020.. |
| [4] | L. Yuchao Shao, "m54366 [VCM] Coding Experiments of End-to-end Compression Network in VCM," June 2020. |
| [5] | C. Hollmann, M. Damghanian and L. Litwic, "m54257 [VCM] Anchor results for semantic segmentation on KITTI dataset," June 2020. |
| [6] | K. Misra, P. Cowan and T. J. a. A. Segall, "m54474 [VCM] Comments on the VCM evaluation framework," June 2020. |
| [7] | S.-P. Wang, E.-C. Ke, C.-C. Lin and C.-L. L. (ITRI), "m54349 [VCM] Anchor generation results for object detection on COCO dataset," June 2020. |
| [8] | N. Le, R. Ghaznavi-Youvalari, A. Zare, H. Zhang, F. Cricri, H. R. Tavakoli, E. Aksu and M. Hannuksela, "m54380 [VCM] VCM anchors," June 2020. |
| [9] | S. Kwak, J. Yun, W.-S. Cheong, J. S. (ETRI), H. Han and H. C. (HNU), "m54412 [VCM] Some considerations and results on VCM anchor generation for object segmentation on Cityscapes dataset," June 2020. |
| [10] | R. Yang, X. Fang, Z. Gao, J. Boyce and H. Yin, "m54339 [VCM] Anchor results on ML Video Test Set – DAVIS 2016 & 2017," June 2020. |

# Appendix A: Anchor Generation Results

## Object Detection

|  |  |  |
| --- | --- | --- |
| **ITRI** [7] | | |
| Network model: Faster R-CNN X101-FPN | | Dataset: COCO 2017 |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP, AP50, AP75, APs, APm, APl * Compression ratio: BPP |
|  | | |
| Test Results: | * Average precision * Large target: APl * Medium target: APm * Small target: APs | |
|  | | |
|  | | |
|  | | |
|  | | |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Scale** | **QP** | **AP** | **AP50** | **AP75** | **APs** | **APm** | **APl** | **BPP** | | **100%** | 22 | 42.2539 | 62.7345 | 46.0454 | 26.2578 | 45.2981 | 54.4273 | 1.6355572 | | 27 | 41.1526 | 61.3616 | 44.802 | 25.4485 | 44.1865 | 53.3624 | 1.017646 | | 32 | 38.3669 | 58.1271 | 41.4496 | 22.6798 | 41.2059 | 50.709 | 0.583934 | | 37 | 32.8522 | 50.9967 | 35.02 | 17.0979 | 35.2142 | 45.46 | 0.3090002 | | 42 | 24.4747 | 39.2073 | 25.8795 | 11.1082 | 26.359 | 36.2924 | 0.1470454 | | 47 | 13.2946 | 22.2228 | 13.5626 | 3.5066 | 13.3497 | 23.2471 | 0.0617982 | | 52 | 5.0019 | 8.678 | 4.9253 | 0.6725 | 3.7845 | 10.5214 | 0.0264967 | | **75%** | 22 | 40.3422 | 60.5121 | 43.6009 | 23.0391 | 43.9257 | 52.9385 | 0.8455161 | | 27 | 38.7834 | 58.5317 | 41.734 | 21.2765 | 42.0859 | 51.8927 | 0.5293182 | | 32 | 35.3652 | 54.1504 | 38.1705 | 19.0389 | 38.4884 | 48.0431 | 0.3091126 | | 37 | 28.463 | 44.8769 | 29.8849 | 13.1841 | 30.7533 | 41.5047 | 0.1648809 | | 42 | 19.3283 | 31.4609 | 20.1164 | 5.9408 | 20.1758 | 32.4532 | 0.0810502 | | 47 | 9.4319 | 15.9271 | 9.4934 | 1.6627 | 8.2877 | 18.5628 | 0.0371622 | | 52 | 3.0709 | 5.4187 | 2.9982 | 0.2151 | 1.8087 | 6.7969 | 0.017438 | | **50%** | 22 | 35.8262 | 54.5561 | 38.2784 | 16.8773 | 39.3136 | 50.6414 | 0.3992588 | | 27 | 33.3388 | 51.6021 | 35.5734 | 15.3319 | 36.3922 | 47.7747 | 0.2556768 | | 32 | 28.4121 | 44.7897 | 29.7963 | 12.5282 | 30.4964 | 42.765 | 0.1515584 | | 37 | 20.8161 | 33.7971 | 21.6769 | 7.2466 | 21.9173 | 34.6354 | 0.0824658 | | 42 | 12.115 | 20.2287 | 12.3801 | 2.4839 | 11.2089 | 22.995 | 0.0423714 | | 47 | 5.0959 | 8.8422 | 5.0844 | 0.469 | 3.5121 | 11.0477 | 0.0206903 | | 52 | 1.5598 | 2.8165 | 1.5139 | 0.0363 | 0.6146 | 3.3515 | 0.0103917 | | **25%** | 22 | 20.6081 | 33.451 | 21.4914 | 4.1636 | 20.8139 | 36.4581 | 0.1221279 | | 27 | 17.8944 | 29.4395 | 18.5176 | 3.5263 | 17.7442 | 32.3513 | 0.0809788 | | 32 | 13.6271 | 22.575 | 14.1926 | 2.0718 | 12.635 | 26.3382 | 0.0499029 | | 37 | 8.2584 | 14.1482 | 8.3805 | 0.8359 | 6.4513 | 17.3047 | 0.0287733 | | 42 | 3.9905 | 7.0315 | 3.8854 | 0.2278 | 2.317 | 9.0308 | 0.0160056 | | 47 | 1.4188 | 2.6146 | 1.4011 | 0.0236 | 0.5079 | 3.182 | 0.0087005 | | 52 | 0.4677 | 0.8657 | 0.4456 | 0.0149 | 0.1351 | 1.011 | 0.0051429 | | original | - | 43.047 | 63.6518 | 46.8884 | 27.205 | 46.0934 | 54.8902 | - | | | |

|  |  |  |
| --- | --- | --- |
| **Nokia** [8] | | |
| Network model: Faster R-CNN X101-FPN | | Dataset: COCO train2017, CityPersons (Eval) |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP [.5:.05:.95], Range: [0,100] * Compression ratio: BPP |
|  | | |
| Test Results: | * Average precision | |
|  | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Resolution** | **QP** | **AP (vtm8.2)** | **BPP (vtm8.2)** | | 25 | 52 | 0.217821782 | 0.001606674 | | 25 | 47 | 1.060218452 | 0.003022804 | | 50 | 52 | 1.251738316 | 0.003640129 | | 75 | 52 | 2.6087042 | 0.006001289 | | 25 | 42 | 3.297143286 | 0.005870285 | | 50 | 47 | 4.067617297 | 0.007295311 | | 100 | 52 | 4.31578773 | 0.008591583 | | 75 | 47 | 7.633100342 | 0.011920967 | | 25 | 37 | 7.81514916 | 0.011045952 | | 50 | 42 | 9.929199551 | 0.014164742 | | 100 | 47 | 10.96856773 | 0.016847031 | | 25 | 32 | 12.68304326 | 0.019959854 | | 75 | 42 | 14.06675963 | 0.022941063 | | 50 | 37 | 15.87992064 | 0.026491043 | | 25 | 27 | 15.93702155 | 0.034098663 | | 100 | 42 | 17.14267054 | 0.032091621 | | 25 | 22 | 17.78900735 | 0.054958305 | | 75 | 37 | 19.07236531 | 0.042192505 | | 50 | 32 | 19.54585627 | 0.048113205 | | 100 | 37 | 20.61614768 | 0.058399925 | | 75 | 32 | 21.74183309 | 0.075827286 | | 50 | 27 | 21.77296789 | 0.083906288 | | 100 | 32 | 22.86530884 | 0.105185928 | | 50 | 22 | 23.43193292 | 0.142293182 | | 75 | 27 | 23.48908098 | 0.133235687 | | 75 | 22 | 24.26911104 | 0.232674042 | | 100 | 27 | 24.32621295 | 0.18898262 | | 100 | 22 | 24.93563888 | 0.344835403 | | BASELINE |  | 25.54793209 | 9.013415 | | | |

## Object Segmentation

|  |  |  |
| --- | --- | --- |
| **ETRI & HNU** [9] | | |
| Network model: R50-FPN-3x COCO  Pre-trained on the Cityscapes dataset is used (initialized from COCO pre-training, then trained on Cityscapes fine annotations only) | | Dataset: Cityscapes |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Range opt: [\* 0 0 \*], [\* 0 1 \*], [\* 1 0 \*], [\* 1 1 \*] * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP * Compression ratio: BPP |
|  | | |
| Test Results: | * Anchor experiment for object segmentation (Cityscape) * Resolution 25%, 50%, 75%, 100% | |
|  | | |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **QP** | **Res.**  **[%]** | = **[\* 0 0 \*]** | | | = **[\* 0 1 \*]** | | | = **[\* 1 0 \*]** | | | | = **[\* 1 1 \*]** | | | | **BPP** | **AP** | **PSNR**  **[dB]** | **BPP** | **AP** | **PSNR**  **[dB]** | **BPP** | **AP** | **PSNR**  **[dB]** | **BPP** | | **AP** | **PSNR**  **[dB]** | | 22 | 25 | 0.05473 | 15.7820 | 32.28 | 0.05473 | 15.4884 | 27.79 | 0.05866 | 16.4603 | 26.64 | 0.05866 | | 16.7739 | 32.79 | | 22 | 50 | 0.14199 | 29.0829 | 36.14 | 0.14199 | 28.3921 | 28.76 | 0.15078 | 29.6022 | 27.44 | 0.15078 | | 30.1879 | 37.10 | | 22 | 75 | 0.23234 | 33.4495 | 37.93 | 0.23234 | 32.6277 | 29.02 | 0.24779 | 32.6997 | 27.66 | 0.24779 | | 32.7472 | 39.27 | | 22 | 100 | 0.34462 | 34.6510 | 39.42 | 0.34462 | 33.8648 | 29.13 | 0.36981 | 34.4493 | 27.83 | 0.36981 | | 34.6058 | 41.07 | | 27 | 25 | 0.03387 | 11.7796 | 31.65 | 0.03387 | 11.7210 | 27.56 | 0.03670 | 12.6900 | 26.47 | 0.03670 | | 12.6756 | 32.14 | | 27 | 50 | 0.08366 | 24.5419 | 35.17 | 0.08366 | 23.9616 | 28.56 | 0.08991 | 25.9446 | 27.30 | 0.08991 | | 26.0047 | 35.99 | | 27 | 75 | 0.13304 | 30.0257 | 36.82 | 0.13304 | 29.7113 | 28.85 | 0.14341 | 30.4809 | 27.54 | 0.14341 | | 30.2268 | 37.90 | | 27 | 100 | 0.18877 | 31.4054 | 38.09 | 0.18877 | 30.5185 | 28.98 | 0.20452 | 32.2868 | 27.73 | 0.20452 | | 32.5538 | 39.33 | | 32 | 25 | 0.01974 | 7.3539 | 30.62 | 0.01974 | 6.4474 | 27.14 | 0.02164 | 8.1445 | 26.16 | 0.02164 | | 7.8131 | 31.10 | | 32 | 50 | 0.04788 | 18.6574 | 33.81 | 0.04788 | 17.7651 | 28.22 | 0.05187 | 19.8431 | 27.05 | 0.05187 | | 19.3413 | 34.50 | | 32 | 75 | 0.07565 | 24.0675 | 35.39 | 0.07565 | 23.3940 | 28.59 | 0.08191 | 25.5255 | 27.36 | 0.08191 | | 24.5363 | 36.24 | | 32 | 100 | 0.10497 | 26.3651 | 36.55 | 0.10497 | 26.0952 | 28.75 | 0.11375 | 28.3342 | 27.57 | 0.11375 | | 27.9029 | 37.47 | | 37 | 25 | 0.01083 | 2.8462 | 29.20 | 0.01083 | 2.7461 | 26.45 | 0.01199 | 4.2937 | 25.64 | 0.01199 | | 3.6257 | 29.66 | | 37 | 50 | 0.02628 | 11.2056 | 32.10 | 0.02628 | 9.7791 | 27.68 | 0.02867 | 12.4545 | 26.65 | 0.02867 | | 11.4918 | 32.67 | | 37 | 75 | 0.04196 | 16.3324 | 33.64 | 0.04196 | 15.6668 | 28.15 | 0.04571 | 17.7395 | 27.04 | 0.04571 | | 17.2819 | 34.30 | | 37 | 100 | 0.05818 | 19.7077 | 34.74 | 0.05818 | 19.1387 | 28.38 | 0.06326 | 21.4306 | 27.30 | 0.06326 | | 21.4169 | 35.43 | | 42 | 25 | 0.00565 | 0.8251 | 27.48 | 0.00565 | 0.7389 | 25.45 | 0.00630 | 1.6717 | 24.87 | 0.00630 | | 1.1388 | 27.93 | | 42 | 50 | 0.01394 | 4.3944 | 30.19 | 0.01394 | 4.2300 | 26.89 | 0.01530 | 6.1293 | 26.06 | 0.01530 | | 5.4339 | 30.68 | | 42 | 75 | 0.02273 | 9.1960 | 31.71 | 0.02273 | 7.5128 | 27.50 | 0.02491 | 10.4364 | 26.58 | 0.02491 | | 9.4581 | 32.26 | | 42 | 100 | 0.03187 | 12.1327 | 32.80 | 0.03187 | 10.9548 | 27.84 | 0.03484 | 14.2091 | 26.91 | 0.03484 | | 13.4861 | 33.36 | | 47 | 25 | 0.00281 | 0.2044 | 25.51 | 0.00281 | 0.1865 | 24.14 | 0.00313 | 0.2872 | 23.76 | 0.00313 | | 0.2628 | 25.93 | | 47 | 50 | 0.00707 | 1.0621 | 28.01 | 0.00707 | 1.3555 | 25.74 | 0.00779 | 2.2251 | 25.14 | 0.00779 | | 1.5390 | 28.44 | | 47 | 75 | 0.01171 | 3.2667 | 29.45 | 0.01171 | 2.9276 | 26.50 | 0.01290 | 4.3832 | 25.81 | 0.01290 | | 3.8061 | 29.92 | | 47 | 100 | 0.01663 | 4.5527 | 30.49 | 0.01663 | 4.1527 | 26.95 | 0.01828 | 6.6726 | 26.25 | 0.01828 | | 5.6201 | 30.95 | | 52 | 25 | 0.00139 | 0.0235 | 23.53 | 0.00139 | 0.0342 | 22.58 | 0.00152 | 0.0350 | 22.43 | 0.00152 | | 0.0412 | 23.87 | | 52 | 50 | 0.00342 | 0.2132 | 25.73 | 0.00342 | 0.2127 | 24.23 | 0.00378 | 0.2971 | 23.94 | 0.00378 | | 0.2627 | 26.11 | | 52 | 75 | 0.00578 | 0.6176 | 27.07 | 0.00578 | 0.5782 | 25.11 | 0.00641 | 1.2569 | 24.72 | 0.00641 | | 0.9080 | 27.46 | | 52 | 100 | 0.00837 | 0.8478 | 28.04 | 0.00837 | 0.6561 | 25.66 | 0.00925 | 1.8011 | 25.26 | 0.00925 | | 1.1679 | 28.43 | | **Average** | | **0.06178** | **13.3782** | **31.91** | **0.06178** | **12.8878** | **27.16** | **0.06660** | **14.3494** | **26.25** | **0.06660** | | **14.0111** | **32.58** | | | |

|  |  |  |
| --- | --- | --- |
| **Foxconn** [2] | | |
| Network model: R50-FPN-3x COCO | | Dataset: COCO 2017 |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP, AP50, AP75, APs, APm, APl * Compression ratio: BPP |
|  | | |
| Test Results: | * Method 2 | |
|  | | |
|  | | |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Resolution** | **QP** | **AP** | **AP50** | **AP75** | **APs** | **APm** | **APl** | | 100% | 22 | 36.5 | 57.859 | 38.844 | 17.834 | 38.793 | 52.404 | | 27 | 35.505 | 56.383 | 37.834 | 17.215 | 37.586 | 51.668 | | 32 | 33.291 | 53.576 | 35.329 | 15.428 | 35.141 | 49.282 | | 37 | 28.815 | 47.34 | 30.354 | 11.487 | 30.075 | 45.059 | | 42 | 21.684 | 37.117 | 21.967 | 7.577 | 22.425 | 37.001 | | 47 | 12.267 | 21.835 | 11.969 | 2.709 | 11.573 | 24.361 | | 52 | 4.632 | 8.697 | 4.319 | 0.363 | 3.32 | 10.853 | | 75% | 22 | 35.163 | 55.758 | 37.515 | 16.243 | 37.23 | 51.928 | | 27 | 33.666 | 53.82 | 35.773 | 14.778 | 35.571 | 50.721 | | 32 | 30.712 | 50.007 | 32.366 | 12.34 | 32.29 | 47.426 | | 37 | 25.1 | 41.782 | 26.23 | 9.597 | 25.863 | 41.385 | | 42 | 17.348 | 30.096 | 17.333 | 4.011 | 17.262 | 32.115 | | 47 | 8.84 | 15.918 | 8.434 | 1.177 | 7.19 | 19.278 | | 52 | 2.933 | 5.537 | 2.69 | 0.226 | 1.667 | 6.942 | | 50% | 22 | 31.171 | 50.392 | 33 | 11.535 | 32.846 | 49.404 | | 27 | 29.272 | 47.886 | 30.938 | 10.597 | 30.178 | 47.153 | | 32 | 24.94 | 41.65 | 25.876 | 8.264 | 25.319 | 42.408 | | 37 | 18.672 | 32.175 | 18.909 | 5.107 | 18.347 | 34.323 | | 42 | 11.21 | 20.281 | 10.525 | 1.62 | 9.73 | 23.672 | | 47 | 4.59 | 8.705 | 4.148 | 0.325 | 2.868 | 11.203 | | 52 | 1.419 | 2.793 | 1.237 | 0.039 | 0.542 | 3.502 | | 25% | 22 | 18.181 | 31.56 | 18.232 | 2.946 | 17.091 | 35.74 | | 27 | 16.055 | 28.15 | 16.042 | 2.473 | 14.598 | 32.029 | | 32 | 12.119 | 21.784 | 11.831 | 1.211 | 10.234 | 25.574 | | 37 | 7.499 | 14.027 | 7.108 | 0.533 | 5.485 | 17.569 | | 42 | 3.741 | 7.155 | 3.341 | 0.26 | 1.928 | 9.406 | | 47 | 1.369 | 2.674 | 1.191 | 0.022 | 0.386 | 3.419 | | 52 | 0.455 | 0.897 | 0.43 | 0.013 | 0.074 | 1.101 | | | |

|  |  |  |
| --- | --- | --- |
| **Ericsson** [5] | | |
| Network model: R50-FPN-3x COCO | | Dataset: KITTI |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * InternalBitDepth=10 * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP * Compression ratio: BPP |
|  | | |
|  | | |
| Test Results: | * Alternative Pipeline 10b | |
|  | | |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Resolution | Uncompressed | QP 22 | QP 27 | QP 32 | QP 37 | QP 42 | QP 47 | QP 52 | | 100% | 84.25% | 84.13% | 83.54% | 82.58% | 80.02% | 74.56% | 61.94% | 37.24% | | 75% | -- | 83.90% | 82.96% | 81.43% | 77.66% | 70.34% | 52.53% | 27.73% | | 50% | -- | 83.30% | 82.22% | 79.33% | 72.44% | 60.39% | 35.86% | 17.50% | | 25% | -- | 80.82% | 78.07% | 72.01% | 58.44% | 34.62% | 17.90% | 15.12% |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 100% | 1.131 | 0.656 | 0.357 | 0.188 | 0.098 | 0.048 | 0.024 | | 75% | 0.696 | 0.416 | 0.234 | 0.126 | 0.066 | 0.032 | 0.016 | | 50% | 0.360 | 0.223 | 0.130 | 0.071 | 0.037 | 0.018 | 0.009 | | 25% | 0.110 | 0.072 | 0.045 | 0.026 | 0.014 | 0.007 | 0.004 | | | |
|  | | |

|  |  |  |
| --- | --- | --- |
| **Nokia** [8] | | |
| Network model: R50-FPN | | Dataset: COCO train2017, Cityscapes (Eval) |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * Scaling resolution: 100%, 75%, 50% and 25% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: AP [.5:.05:.95], Range: [0,100] * Compression ratio: BPP |
|  | | |
| Test Results: | * Average precision | |
|  | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Resolution** | **QP** | **AP (vtm8.2)** | **BPP (vtm8.2)** | | 25 | 52 | 0.253440258 | 0.001606674 | | 50 | 52 | 0.260041895 | 0.003640129 | | 25 | 47 | 0.410809252 | 0.003022804 | | 75 | 52 | 0.4752796 | 0.006001289 | | 100 | 52 | 0.847828181 | 0.008591583 | | 50 | 47 | 1.414990844 | 0.007295311 | | 25 | 42 | 1.536290758 | 0.005870285 | | 75 | 47 | 3.272933999 | 0.011920967 | | 100 | 47 | 4.552687183 | 0.016847031 | | 25 | 37 | 4.797686711 | 0.011045952 | | 50 | 42 | 5.137436638 | 0.014164742 | | 75 | 42 | 9.088325461 | 0.022941063 | | 25 | 32 | 10.50516903 | 0.019959854 | | 50 | 37 | 10.90742777 | 0.026491043 | | 100 | 42 | 12.13270528 | 0.032091621 | | 75 | 37 | 16.38746211 | 0.042192505 | | 25 | 27 | 18.17095985 | 0.034098663 | | 50 | 32 | 19.57606609 | 0.048113205 | | 100 | 37 | 19.70766409 | 0.058399925 | | 25 | 22 | 24.02271044 | 0.054958305 | | 75 | 32 | 24.59459364 | 0.075827286 | | 50 | 27 | 25.77488773 | 0.083906288 | | 100 | 32 | 26.3651416 | 0.105185928 | | 75 | 27 | 29.96757122 | 0.133235687 | | 100 | 27 | 31.60014316 | 0.18898262 | | 50 | 22 | 31.70485412 | 0.142293182 | | 75 | 22 | 33.43511414 | 0.232674042 | | 100 | 22 | 34.65097283 | 0.344835403 | | BASELINE |  | 36.48087673 | 9.013415 | | | |

## Video Segmentation

|  |  |  |
| --- | --- | --- |
| **Intel** [10] | | |
| Network model: PReMVOS | | Dataset: DAVIS 2016, DAVIS 2017 |
| Test Condition: | | * Format conversion: FFmpeg 4.2.2 * Codec: VTM 8.2 * CTC Profile for Video: RA 8-bits * Scaling resolution: 100% * QPs: 22, 27, 32, 37, 42, 47, 52 * Average precision: J&F-Mean * Compression ratio: BPP |
|  | | |
| Test Results: | * J&F-Mean DAVIS 2016 * J&F-Mean DAVIS 2017 | |
| A close up of a map  Description automatically generated | | |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | J&F-Mean | J-Mean | J-Recall | J-Decay | F-Mean | F-Recall | F-Decay | | Original | 0.865 | 0.857 | 0.96 | 0.068 | 0.873 | 0.954 | 0.063 | | QP22 | 0.866 | 0.856 | 0.956 | 0.082 | 0.876 | 0.958 | 0.073 | | QP27 | 0.863 | 0.853 | 0.948 | 0.062 | 0.873 | 0.951 | 0.066 | | QP32 | 0.856 | 0.848 | 0.947 | 0.064 | 0.865 | 0.944 | 0.066 | | QP37 | 0.848 | 0.841 | 0.942 | 0.057 | 0.856 | 0.931 | 0.059 | | QP42 | 0.821 | 0.815 | 0.929 | 0.052 | 0.827 | 0.909 | 0.066 | | QP47 | 0.766 | 0.762 | 0.863 | 0.082 | 0.769 | 0.868 | 0.101 | | QP52 | 0.677 | 0.682 | 0.798 | 0.1 | 0.671 | 0.78 | 0.123 |   DAVIS 2016 Overall Performance Indicators | | |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | Original | | QP22 | | QP27 | | QP32 | | QP37 | | QP42 | | QP47 | | QP52 | | | Sequence | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | | blackswan\_1 | 0.955 | 0.98 | 0.955 | 0.979 | 0.953 | 0.978 | 0.951 | 0.974 | 0.947 | 0.972 | 0.943 | 0.963 | 0.94 | 0.968 | 0.919 | 0.945 | | bmx-trees\_1 | 0.548 | 0.746 | 0.565 | 0.773 | 0.532 | 0.732 | 0.499 | 0.71 | 0.477 | 0.685 | 0.404 | 0.592 | 0.312 | 0.472 | 0.201 | 0.34 | | breakdance\_1 | 0.814 | 0.838 | 0.83 | 0.855 | 0.799 | 0.838 | 0.824 | 0.85 | 0.789 | 0.824 | 0.76 | 0.779 | 0.625 | 0.643 | 0.533 | 0.55 | | camel\_1 | 0.945 | 0.985 | 0.943 | 0.986 | 0.942 | 0.986 | 0.94 | 0.985 | 0.936 | 0.983 | 0.93 | 0.979 | 0.913 | 0.966 | 0.872 | 0.918 | | car-roundabout\_1 | 0.977 | 0.937 | 0.977 | 0.938 | 0.978 | 0.94 | 0.977 | 0.94 | 0.976 | 0.937 | 0.974 | 0.936 | 0.97 | 0.931 | 0.963 | 0.927 | | car-shadow\_1 | 0.97 | 0.994 | 0.969 | 0.994 | 0.969 | 0.995 | 0.968 | 0.995 | 0.965 | 0.994 | 0.963 | 0.994 | 0.955 | 0.99 | 0.946 | 0.99 | | cows\_1 | 0.934 | 0.953 | 0.934 | 0.954 | 0.934 | 0.954 | 0.933 | 0.953 | 0.932 | 0.952 | 0.931 | 0.951 | 0.923 | 0.941 | 0.894 | 0.897 | | dance-twirl\_1 | 0.787 | 0.797 | 0.797 | 0.809 | 0.797 | 0.813 | 0.797 | 0.807 | 0.797 | 0.813 | 0.787 | 0.801 | 0.753 | 0.77 | 0.688 | 0.678 | | dog\_1 | 0.941 | 0.965 | 0.94 | 0.963 | 0.937 | 0.958 | 0.931 | 0.945 | 0.92 | 0.924 | 0.901 | 0.881 | 0.859 | 0.79 | 0.796 | 0.669 | | drift-chicane\_1 | 0.925 | 0.97 | 0.922 | 0.973 | 0.924 | 0.973 | 0.913 | 0.966 | 0.867 | 0.935 | 0.876 | 0.95 | 0.813 | 0.904 | 0.451 | 0.499 | | drift-straight\_1 | 0.945 | 0.942 | 0.942 | 0.934 | 0.938 | 0.927 | 0.937 | 0.924 | 0.936 | 0.923 | 0.929 | 0.908 | 0.908 | 0.869 | 0.857 | 0.764 | | goat\_1 | 0.893 | 0.904 | 0.891 | 0.901 | 0.89 | 0.899 | 0.884 | 0.886 | 0.881 | 0.882 | 0.868 | 0.863 | 0.848 | 0.817 | 0.781 | 0.706 | | horsejump-high\_1 | 0.868 | 0.905 | 0.869 | 0.908 | 0.867 | 0.91 | 0.864 | 0.914 | 0.858 | 0.908 | 0.848 | 0.898 | 0.829 | 0.871 | 0.773 | 0.831 | | kite-surf\_1 | 0.641 | 0.765 | 0.652 | 0.793 | 0.602 | 0.783 | 0.623 | 0.782 | 0.658 | 0.794 | 0.666 | 0.789 | 0.423 | 0.666 | 0.598 | 0.66 | | libby\_1 | 0.853 | 0.941 | 0.795 | 0.911 | 0.795 | 0.911 | 0.774 | 0.902 | 0.78 | 0.907 | 0.625 | 0.763 | 0.442 | 0.56 | 0.321 | 0.402 | | motocross-jump\_1 | 0.781 | 0.718 | 0.774 | 0.702 | 0.839 | 0.727 | 0.832 | 0.714 | 0.778 | 0.678 | 0.749 | 0.641 | 0.687 | 0.568 | 0.492 | 0.437 | | paragliding-launch\_1 | 0.619 | 0.443 | 0.625 | 0.47 | 0.623 | 0.455 | 0.625 | 0.41 | 0.605 | 0.367 | 0.602 | 0.279 | 0.606 | 0.204 | 0.555 | 0.193 | | parkour\_1 | 0.926 | 0.98 | 0.923 | 0.979 | 0.924 | 0.978 | 0.919 | 0.973 | 0.915 | 0.971 | 0.898 | 0.959 | 0.844 | 0.902 | 0.584 | 0.639 | | scooter-black\_1 | 0.904 | 0.806 | 0.901 | 0.813 | 0.905 | 0.804 | 0.902 | 0.802 | 0.898 | 0.799 | 0.884 | 0.786 | 0.859 | 0.768 | 0.754 | 0.658 | | soapbox\_1 | 0.905 | 0.887 | 0.912 | 0.894 | 0.912 | 0.893 | 0.858 | 0.872 | 0.896 | 0.88 | 0.771 | 0.836 | 0.73 | 0.786 | 0.654 | 0.725 |   DAVIS 2016 Per-Sequence, J-Mean & F-Mean | | |
|  | | |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | J&F-Mean | J-Mean | J-Recall | J-Decay | F-Mean | F-Recall | F-Decay | | Original | 0.698 | 0.667 | 0.739 | 0.176 | 0.728 | 0.791 | 0.207 | | QP22 | 0.718 | 0.686 | 0.758 | 0.13 | 0.75 | 0.809 | 0.17 | | QP27 | 0.697 | 0.666 | 0.735 | 0.156 | 0.729 | 0.786 | 0.195 | | QP32 | 0.693 | 0.659 | 0.725 | 0.166 | 0.727 | 0.779 | 0.198 | | QP37 | 0.688 | 0.658 | 0.736 | 0.116 | 0.717 | 0.786 | 0.168 | | QP42 | 0.659 | 0.628 | 0.702 | 0.165 | 0.69 | 0.755 | 0.198 | | QP47 | 0.592 | 0.564 | 0.639 | 0.215 | 0.62 | 0.698 | 0.242 | | QP52 | 0.532 | 0.508 | 0.588 | 0.166 | 0.556 | 0.64 | 0.218 |   DAVIS 2017 Overall Performance Indicators | | |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | Original | | QP22 | | QP27 | | QP32 | | QP37 | | QP42 | | QP47 | | QP52 | | | Sequence | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | J-Mean | F-Mean | | bike-packing\_1 | 0.46 | 0.677 | 0.494 | 0.668 | 0.545 | 0.719 | 0.408 | 0.545 | 0.287 | 0.357 | 0.515 | 0.683 | 0.338 | 0.619 | 0.232 | 0.502 | | bike-packing\_2 | 0.609 | 0.544 | 0.671 | 0.602 | 0.598 | 0.57 | 0.557 | 0.593 | 0.515 | 0.564 | 0.603 | 0.561 | 0.421 | 0.492 | 0.412 | 0.514 | | blackswan\_1 | 0.955 | 0.98 | 0.955 | 0.979 | 0.953 | 0.978 | 0.951 | 0.974 | 0.947 | 0.972 | 0.943 | 0.963 | 0.94 | 0.968 | 0.919 | 0.945 | | bmx-trees\_1 | 0.338 | 0.606 | 0.345 | 0.638 | 0.323 | 0.598 | 0.3 | 0.563 | 0.283 | 0.552 | 0.231 | 0.448 | 0.14 | 0.293 | 0.046 | 0.136 | | bmx-trees\_2 | 0.757 | 0.93 | 0.757 | 0.931 | 0.736 | 0.919 | 0.732 | 0.908 | 0.704 | 0.88 | 0.671 | 0.859 | 0.593 | 0.77 | 0.42 | 0.555 | | breakdance\_1 | 0.814 | 0.838 | 0.83 | 0.855 | 0.799 | 0.838 | 0.824 | 0.85 | 0.789 | 0.824 | 0.76 | 0.779 | 0.625 | 0.643 | 0.533 | 0.55 | | camel\_1 | 0.945 | 0.985 | 0.943 | 0.986 | 0.942 | 0.986 | 0.94 | 0.985 | 0.936 | 0.983 | 0.93 | 0.979 | 0.913 | 0.966 | 0.872 | 0.918 | | car-roundabout\_1 | 0.977 | 0.937 | 0.977 | 0.938 | 0.978 | 0.94 | 0.977 | 0.94 | 0.976 | 0.937 | 0.974 | 0.936 | 0.97 | 0.931 | 0.963 | 0.927 | | car-shadow\_1 | 0.97 | 0.994 | 0.969 | 0.994 | 0.969 | 0.995 | 0.968 | 0.995 | 0.965 | 0.994 | 0.963 | 0.994 | 0.955 | 0.99 | 0.946 | 0.99 | | cows\_1 | 0.934 | 0.953 | 0.934 | 0.954 | 0.934 | 0.954 | 0.933 | 0.953 | 0.932 | 0.952 | 0.931 | 0.951 | 0.923 | 0.941 | 0.894 | 0.897 | | dance-twirl\_1 | 0.787 | 0.797 | 0.797 | 0.809 | 0.797 | 0.813 | 0.797 | 0.807 | 0.797 | 0.813 | 0.787 | 0.801 | 0.753 | 0.77 | 0.688 | 0.678 | | dog\_1 | 0.941 | 0.965 | 0.94 | 0.963 | 0.937 | 0.958 | 0.931 | 0.945 | 0.92 | 0.924 | 0.901 | 0.881 | 0.859 | 0.79 | 0.796 | 0.669 | | dogs-jump\_1 | 0.835 | 0.941 | 0.859 | 0.944 | 0.868 | 0.952 | 0.869 | 0.954 | 0.856 | 0.946 | 0.837 | 0.911 | 0.084 | 0.127 | 0.447 | 0.545 | | dogs-jump\_2 | 0.915 | 0.967 | 0.912 | 0.965 | 0.914 | 0.968 | 0.91 | 0.967 | 0.911 | 0.969 | 0.9 | 0.958 | 0.12 | 0.129 | 0.47 | 0.486 | | dogs-jump\_3 | 0.91 | 0.949 | 0.901 | 0.952 | 0.914 | 0.967 | 0.909 | 0.961 | 0.907 | 0.955 | 0.892 | 0.94 | 0.877 | 0.948 | 0.839 | 0.918 | | drift-chicane\_1 | 0.925 | 0.97 | 0.922 | 0.973 | 0.924 | 0.973 | 0.913 | 0.966 | 0.867 | 0.935 | 0.876 | 0.95 | 0.813 | 0.904 | 0.451 | 0.499 | | drift-straight\_1 | 0.945 | 0.942 | 0.942 | 0.934 | 0.938 | 0.927 | 0.937 | 0.924 | 0.936 | 0.923 | 0.929 | 0.908 | 0.908 | 0.869 | 0.857 | 0.764 | | goat\_1 | 0.893 | 0.904 | 0.891 | 0.901 | 0.89 | 0.899 | 0.884 | 0.886 | 0.881 | 0.882 | 0.868 | 0.863 | 0.848 | 0.817 | 0.781 | 0.706 | | gold-fish\_1 | 0.858 | 0.87 | 0.856 | 0.868 | 0.848 | 0.857 | 0.84 | 0.852 | 0.785 | 0.784 | 0.736 | 0.741 | 0.716 | 0.706 | 0.694 | 0.677 | | gold-fish\_2 | 0.376 | 0.55 | 0.517 | 0.715 | 0.495 | 0.571 | 0.594 | 0.723 | 0.614 | 0.771 | 0.58 | 0.754 | 0.559 | 0.656 | 0.457 | 0.511 | | gold-fish\_3 | 0.793 | 0.846 | 0.807 | 0.852 | 0.798 | 0.839 | 0.801 | 0.844 | 0.799 | 0.834 | 0.782 | 0.831 | 0.775 | 0.809 | 0.703 | 0.752 | | gold-fish\_4 | 0.902 | 0.931 | 0.897 | 0.927 | 0.888 | 0.92 | 0.877 | 0.907 | 0.882 | 0.914 | 0.888 | 0.923 | 0.869 | 0.904 | 0.847 | 0.891 | | gold-fish\_5 | 0.885 | 0.859 | 0.886 | 0.856 | 0.893 | 0.868 | 0.892 | 0.866 | 0.884 | 0.862 | 0.868 | 0.833 | 0.834 | 0.782 | 0.764 | 0.679 | | horsejump-high\_1 | 0.841 | 0.933 | 0.841 | 0.934 | 0.836 | 0.935 | 0.831 | 0.933 | 0.824 | 0.926 | 0.813 | 0.912 | 0.789 | 0.885 | 0.739 | 0.834 | | horsejump-high\_2 | 0.851 | 0.987 | 0.846 | 0.988 | 0.836 | 0.988 | 0.822 | 0.982 | 0.806 | 0.975 | 0.778 | 0.963 | 0.744 | 0.947 | 0.645 | 0.875 | | india\_1 | 0.679 | 0.665 | 0.688 | 0.673 | 0.452 | 0.484 | 0.68 | 0.66 | 0.671 | 0.653 | 0.653 | 0.625 | 0.621 | 0.584 | 0.555 | 0.489 | | india\_2 | 0.418 | 0.445 | 0.417 | 0.44 | 0.417 | 0.437 | 0.482 | 0.482 | 0.406 | 0.414 | 0.462 | 0.46 | 0.443 | 0.42 | 0.351 | 0.341 | | india\_3 | 0.387 | 0.393 | 0.388 | 0.389 | 0.631 | 0.611 | 0.388 | 0.433 | 0.384 | 0.388 | 0.381 | 0.447 | 0.363 | 0.412 | 0.369 | 0.348 | | judo\_1 | 0.847 | 0.884 | 0.819 | 0.87 | 0.327 | 0.336 | 0.83 | 0.872 | 0.842 | 0.882 | 0.811 | 0.853 | 0.364 | 0.381 | 0.734 | 0.761 | | judo\_2 | 0.316 | 0.334 | 0.315 | 0.334 | 0.309 | 0.332 | 0.311 | 0.33 | 0.318 | 0.334 | 0.318 | 0.347 | 0.347 | 0.367 | 0.143 | 0.145 | | kite-surf\_1 | 0.238 | 0.712 | 0.237 | 0.737 | 0.237 | 0.743 | 0.236 | 0.735 | 0.222 | 0.721 | 0.232 | 0.713 | 0.219 | 0.668 | 0.177 | 0.541 | | kite-surf\_2 | 0.026 | 0.094 | 0.106 | 0.247 | 0.004 | 0.049 | 0.067 | 0.096 | 0.013 | 0.083 | 0.004 | 0.031 | 0.015 | 0.087 | 0.02 | 0.069 | | kite-surf\_3 | 0.714 | 0.934 | 0.711 | 0.937 | 0.699 | 0.925 | 0.691 | 0.924 | 0.694 | 0.927 | 0.688 | 0.916 | 0.661 | 0.893 | 0.628 | 0.841 | | lab-coat\_1 | 0 | 0.011 | 0.001 | 0.015 | 0 | 0.014 | 0.012 | 0.031 | 0 | 0.012 | 0.001 | 0.015 | 0 | 0 | 0 | 0 | | lab-coat\_2 | 0 | 0 | 0 | 0 | 0.014 | 0.022 | 0.015 | 0.022 | 0.013 | 0.069 | 0.001 | 0.034 | 0 | 0 | 0.004 | 0.022 | | lab-coat\_3 | 0.816 | 0.745 | 0.836 | 0.765 | 0.867 | 0.776 | 0.744 | 0.708 | 0.833 | 0.786 | 0.761 | 0.74 | 0.804 | 0.738 | 0.829 | 0.791 | | lab-coat\_4 | 0.687 | 0.625 | 0.75 | 0.659 | 0.723 | 0.66 | 0.766 | 0.667 | 0.848 | 0.754 | 0.808 | 0.677 | 0.741 | 0.644 | 0.328 | 0.388 | | lab-coat\_5 | 0.859 | 0.8 | 0.862 | 0.803 | 0.875 | 0.826 | 0.879 | 0.839 | 0.884 | 0.848 | 0.865 | 0.818 | 0.864 | 0.813 | 0.787 | 0.705 | | libby\_1 | 0.853 | 0.941 | 0.795 | 0.911 | 0.795 | 0.911 | 0.774 | 0.902 | 0.78 | 0.907 | 0.625 | 0.763 | 0.442 | 0.56 | 0.321 | 0.402 | | loading\_1 | 0.956 | 0.964 | 0.956 | 0.964 | 0.955 | 0.964 | 0.954 | 0.962 | 0.952 | 0.961 | 0.951 | 0.961 | 0.947 | 0.953 | 0.923 | 0.909 | | loading\_2 | 0.063 | 0.162 | 0.053 | 0.125 | 0.035 | 0.195 | 0.1 | 0.271 | 0.043 | 0.121 | 0.122 | 0.173 | 0.135 | 0.168 | 0.116 | 0.153 | | loading\_3 | 0.75 | 0.775 | 0.752 | 0.775 | 0.753 | 0.776 | 0.742 | 0.759 | 0.728 | 0.746 | 0.692 | 0.731 | 0.716 | 0.744 | 0.685 | 0.714 | | mbike-trick\_1 | 0.854 | 0.922 | 0.846 | 0.913 | 0.846 | 0.915 | 0.836 | 0.905 | 0.818 | 0.895 | 0.79 | 0.867 | 0.7 | 0.785 | 0.515 | 0.665 | | mbike-trick\_2 | 0.858 | 0.86 | 0.849 | 0.857 | 0.846 | 0.848 | 0.837 | 0.843 | 0.816 | 0.825 | 0.792 | 0.802 | 0.748 | 0.761 | 0.366 | 0.458 | | motocross-jump\_1 | 0.375 | 0.497 | 0.346 | 0.466 | 0.354 | 0.496 | 0.33 | 0.465 | 0.551 | 0.595 | 0.338 | 0.475 | 0.31 | 0.415 | 0.089 | 0.201 | | motocross-jump\_2 | 0.383 | 0.426 | 0.368 | 0.411 | 0.377 | 0.46 | 0.392 | 0.475 | 0.663 | 0.636 | 0.302 | 0.377 | 0.265 | 0.317 | 0.462 | 0.459 | | paragliding-launch\_1 | 0.112 | 0.163 | 0.104 | 0.236 | 0.086 | 0.18 | 0.082 | 0.161 | 0.062 | 0.102 | 0.064 | 0.125 | 0.025 | 0.084 | 0 | 0.039 | | paragliding-launch\_2 | 0.361 | 0.737 | 0.356 | 0.732 | 0.358 | 0.73 | 0.354 | 0.727 | 0.358 | 0.721 | 0.351 | 0.71 | 0.328 | 0.676 | 0.294 | 0.591 | | paragliding-launch\_3 | 0.088 | 0.331 | 0.091 | 0.33 | 0.091 | 0.335 | 0.087 | 0.312 | 0.053 | 0.175 | 0.052 | 0.161 | 0.017 | 0.046 | 0.008 | 0.036 | | parkour\_1 | 0.926 | 0.98 | 0.923 | 0.979 | 0.924 | 0.978 | 0.919 | 0.973 | 0.915 | 0.971 | 0.898 | 0.959 | 0.844 | 0.902 | 0.584 | 0.639 | | pigs\_1 | 0.873 | 0.904 | 0.902 | 0.929 | 0.871 | 0.896 | 0.891 | 0.914 | 0.874 | 0.896 | 0.857 | 0.862 | 0.803 | 0.782 | 0.749 | 0.67 | | pigs\_2 | 0.516 | 0.547 | 0.534 | 0.568 | 0.514 | 0.55 | 0.463 | 0.489 | 0.451 | 0.48 | 0.455 | 0.479 | 0.449 | 0.533 | 0.452 | 0.543 | | pigs\_3 | 0.935 | 0.903 | 0.929 | 0.89 | 0.916 | 0.877 | 0.927 | 0.886 | 0.904 | 0.85 | 0.904 | 0.852 | 0.905 | 0.849 | 0.897 | 0.826 | | scooter-black\_1 | 0.854 | 0.961 | 0.846 | 0.951 | 0.84 | 0.947 | 0.841 | 0.952 | 0.823 | 0.945 | 0.816 | 0.937 | 0.8 | 0.932 | 0.483 | 0.588 | | scooter-black\_2 | 0.869 | 0.847 | 0.862 | 0.846 | 0.868 | 0.842 | 0.862 | 0.842 | 0.854 | 0.833 | 0.841 | 0.814 | 0.816 | 0.798 | 0.776 | 0.752 | | shooting\_1 | 0.112 | 0.157 | 0.484 | 0.557 | 0.067 | 0.182 | 0.043 | 0.148 | 0.065 | 0.162 | 0.043 | 0.086 | 0.027 | 0.067 | 0.021 | 0.063 | | shooting\_2 | 0.143 | 0.215 | 0.624 | 0.636 | 0.556 | 0.574 | 0.09 | 0.233 | 0.116 | 0.176 | 0.13 | 0.283 | 0.092 | 0.278 | 0.112 | 0.143 | | shooting\_3 | 0.84 | 0.963 | 0.852 | 0.963 | 0.883 | 0.97 | 0.714 | 0.839 | 0.686 | 0.8 | 0.012 | 0.122 | 0.006 | 0.09 | 0.031 | 0.123 | | soapbox\_1 | 0.847 | 0.848 | 0.856 | 0.855 | 0.857 | 0.86 | 0.787 | 0.818 | 0.842 | 0.853 | 0.672 | 0.746 | 0.654 | 0.72 | 0.43 | 0.531 | | soapbox\_2 | 0.809 | 0.874 | 0.818 | 0.88 | 0.818 | 0.88 | 0.817 | 0.881 | 0.798 | 0.863 | 0.794 | 0.853 | 0.769 | 0.813 | 0.617 | 0.745 | | soapbox\_3 | 0.893 | 0.961 | 0.89 | 0.958 | 0.886 | 0.96 | 0.885 | 0.961 | 0.869 | 0.95 | 0.85 | 0.94 | 0.809 | 0.906 | 0.71 | 0.812 |   DAVIS 2017 Per-Sequence, J-Mean & F-Mean | | |

# Appendix B: Anchor Runtime Measurement

Runtime as a proxy for complexity in JVET / VVC standardization to be reported for anchors for Machine vision task and Human vision task. The proposed runtime methods for machine and human visions are:

* Machine vision tasks
  + EncTm: Time needed to convert RGB input to bitstream
  + DecTm: Time needed to convert bitstream to inference output
* Human vision
  + Encv: time needed to convert RGB input to bitstream
  + DecTv: Time needed to convert bitstream to YUV

The runtime methods for machine and human visions are shown in Figure 5 and Figure 6, respectively.

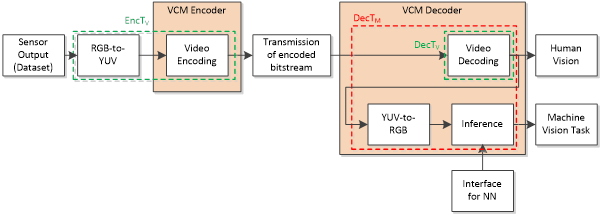


Figure . *Runtime measurements to be made for anchors*

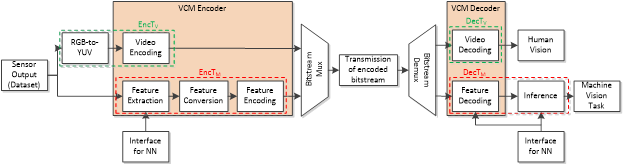


Figure . *Ilustration of runtime measurements to be made for proposed technolog*