INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 2
MPEG TECHNICAL REQUIREMENTS

ISO/IEC JTC 1/SC 29/WG 4 m62360
Online – January 2023

Title: [VCM] Cross-check of CE1.3 (m62009)

Source: Poznań University of Technology
Authors: Marek Domański, Olgierd Stankiewicz, Sławomir Maćkowiak, Tomasz Grajek, Maciej Wawrzyniak, Jakub Stankowski, Sławomir Różek, Dominik Cywiński, Jakub Szekiełda, Jakub Siejak

Abstract

This document describes the procedure and results of crosschecking Core Experiment proposal CE1.3 [1].

1. Introduction

We have performed crosscheck with the use of machines with the following specification:
Processor: Intel Core i7-5820K
RAM: 64 GB
GPU: nVidia RTX 3080.
Operating System: Ubuntu 18.04 (updated)
We had installed Proponent's software without any issues. On the other hand, running attached encoding and decoding scripts has caused minor problems. Additionally, our spare workstations for encoding and decoding were not good enough to work on multiple datasets and/or QPs efficiently at once. All those problems resulted in this document, in which we can only provide our final results for datasets:

- FLIR,
- TVD-images,
- SFU.

It is worth mentioning that Proponent in m62009 provided results only for these datasets as well.

2. Results

2.1. Object Detection

Table. 1. Object detection results on FLIR and TVD datasets

			CE1.3.		OUR cross-check		Difference	
Scale	Dataset	QP	BPP	mAP	BPP	mAP	BPP	mAP
100%	FLIR	22	0,388	39,052	0,387	38,173	0,000	0,879
		27	0,256	38,364	0,256	37,380	0,000	0,984
		32	0,108	35,213	0,108	35,781	0,000	-0,569
		37	0,059	31,232	0,059	30,264	0,000	0,968
		42	0,033	20,734	0,033	20,432	0,000	0,302
		47	0,021	11,032	0,021	10,434	0,000	0,598
	TVD	22	0,153	53,465	0,154	54,038	0,000	-0,573
		27	0,090	50,647	0,090	50,761	0,000	-0,114
		32	0,052	46,730	0,052	46,694	0,000	0,036
		37	0,029	42,832	0,029	42,872	0,000	-0,040
		42	0,016	32,618	0,016	31,822	0,000	0,797
		47	0,010	18,549	0,010	18,865	0,000	-0,315

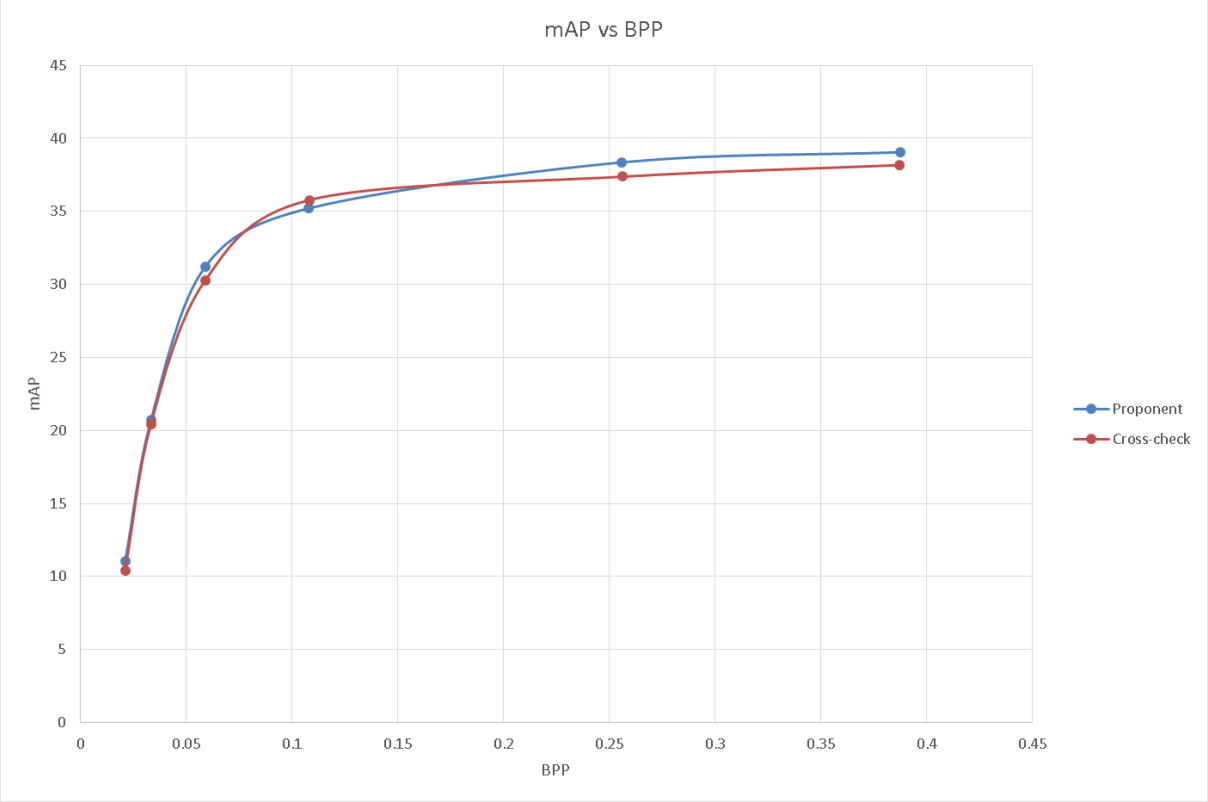

Fig. 1. Object detection result on FLIR dataset

Fig. 2. Exemplary decoded image FLIR_08931(qp 37)

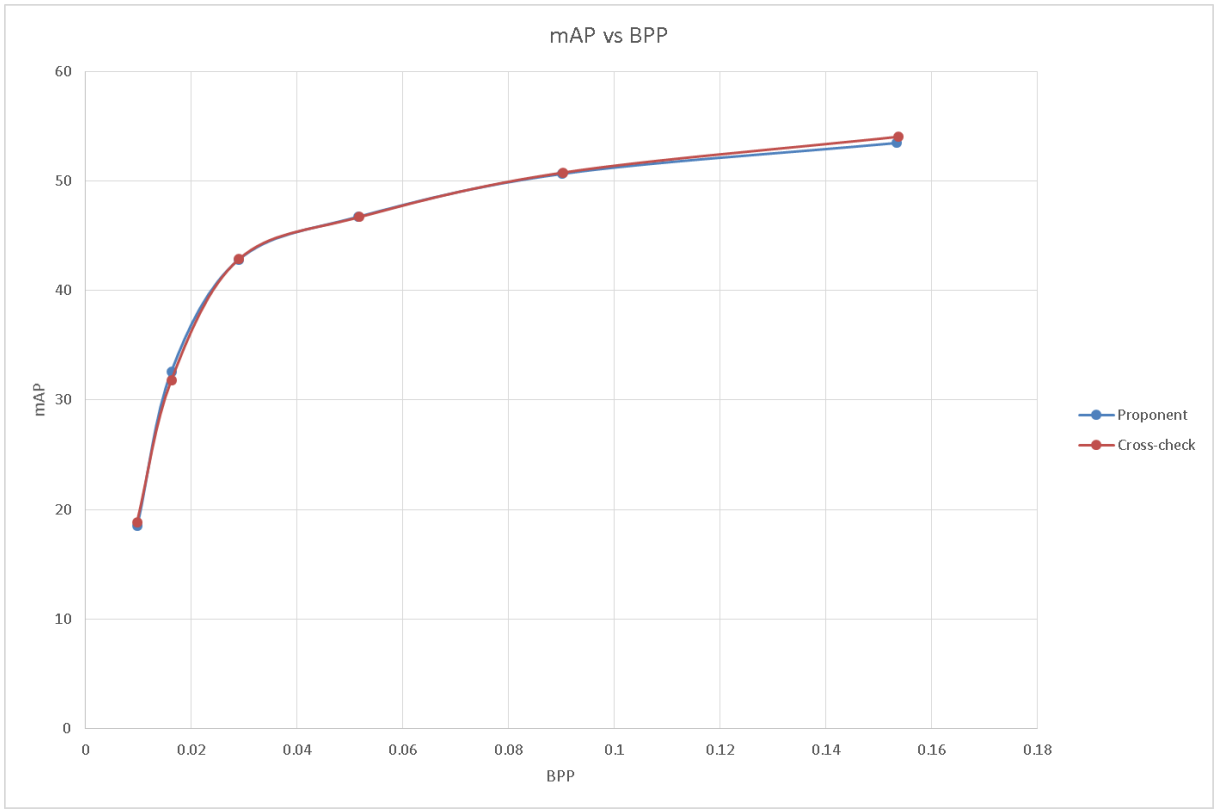

Fig. 3. Object detection result on TVD dataset

Fig. 4. Exemplary decoded image TVD 000069 (qp 47)

Table. 2. Object detection results on SFU datasets

			CE.1.3.		OUR Cross-check		Difference	
	Sequence	QP	kbps	mAP	kbps	mAP	kbps	mAP
Class A 100%	Traffic_ 2560x1600_30_val	37	1351,96	41,04	1351,27	41,09	0,69	-0,05
		42	759,19	40,71	757,12	41,10	2,07	-0,38
		47	422,80	39,91	422,79	39,64	0,01	0,27
		52	219,76	39,28	219,05	38,23	0,71	1,06
		57	110,36	32,54	109,72	33,71	0,65	-1,17
		62	59,02	19,11	59,10	12,21	-0,08	6,90
Class B 100%	Kimono_ 1920x1080_24_val	32	150,54	75,15	150,54	75,15	0,00	0,00
		37	84,33	76,03	84,33	76,01	0,00	0,02
		42	47,83	73,11	47,83	73,56	0,00	-0,45
		47	31,41	73,93	31,41	73,84	0,00	0,08
		52	19,62	77,29	19,62	77,04	0,00	0,25
		57	13,35	37,81	13,35	37,80	0,00	0,01
	ParkScene_ 1920x1080_24_val	32	868,98	54,57	864,71	55,71	4,27	-1,13
		37	448,05	47,26	446,71	52,11	1,34	-4,85
		42	227,10	45,39	226,63	43,52	0,47	1,87
		47	115,30	36,19	114,91	35,29	0,39	0,90
		52	53,66	24,53	53,75	27,15	-0,09	-2,62
		57	27,37	8,98	26,93	8,31	0,44	0,68
	Cactus_ 1920x1080_50_val	32	1993,45	72,17	2023,78	71,57	-30,33	0,60
		37	1033,68	69,92	1046,64	68,67	-12,95	1,24
		42	532,21	69,54	536,06	69,34	-3,85	0,20
		47	272,33	71,09	273,65	69,80	-1,31	1,29
		52	133,00	63,09	133,81	62,31	-0,81	0,77
		57	69,34	29,42	69,73	30,54	-0,38	-1,12
	BasketballDrive_ 1920x1080_50_val	32	2185,92	42,31	2166,03	41,49	19,89	0,81
		37	1149,90	41,69	1144,51	40,36	5,39	1,34
		42	604,14	41,49	604,16	41,13	-0,02	0,36
		47	309,93	38,09	311,84	37,30	-1,91	0,78
		52	156,12	27,67	157,12	26,31	-0,99	1,36
		57	83,72	12,22	83,96	13,86	-0,24	-1,64
	BQTerrace_ 1920x1080_60_val	32	1802,18	42,12	1820,47	42,65	-18,29	-0,53
		37	911,99	42,47	922,06	42,40	-10,07	0,07
		42	478,02	42,36	483,23	42,09	-5,21	0,28
		47	250,20	39,62	252,83	40,07	-2,63	-0,45
		52	125,26	28,90	126,89	31,54	-1,63	-2,64
		57	60,97	14,82	61,13	14,47	-0,16	0,35

Class C 100%	BasketballDrill_ 832x480_50_val	27	1455,05	25,61	1455,05	25,62	0,00	-0,01
		32	735,26	22,64	735,26	22,55	0,00	0,09
		37	385,58	17,40	385,58	17,39	0,00	0,00
		42	205,32	12,51	205,32	12,49	0,00	0,01
		47	104,77	7,06	104,77	7,07	0,00	0,00
		52	52,11	3,16	52,11	3,15	0,00	0,00
	BQMall_ 832x480_60_val	27	1853,92	44,55	1853,92	44,62	0,00	-0,07
		32	950,21	43,82	950,21	43,84	0,00	-0,03
		37	511,28	40,38	511,28	40,38	0,00	0,01
		42	276,43	36,10	276,43	36,11	0,00	-0,01
		47	145,85	32,88	145,85	32,89	0,00	-0,01
		52	75,61	23,93	75,61	23,91	0,00	0,01
	PartyScene_ 832x480_50_val	27	1831,73	68,16	1830,99	68,46	0,74	-0,30
		32	892,09	64,26	891,27	67,31	0,82	-3,05
		37	449,69	61,89	448,05	59,19	1,65	2,70
		42	219,10	55,44	218,18	53,39	0,92	2,06
		47	102,71	26,88	102,71	26,86	0,00	0,02
		52	45,65	19,43	45,65	19,47	0,00	-0,04
	RaceHorses_ 832x480_30_val	27	1169,40	45,44	1169,40	45,43	0,00	0,01
		32	598,19	43,02	598,19	43,04	0,00	-0,02
		37	322,35	38,12	322,35	38,08	0,00	0,04
		42	172,33	34,32	172,33	34,31	0,00	0,01
		47	90,40	24,47	90,40	24,43	0,00	0,05
		52	47,07	12,31	47,07	12,37	0,00	-0,06
Class D 100%	BasketballPass_ 416x240_50_val	22	1256,28	27,41	1256,28	27,37	0,00	0,04
		27	644,02	24,95	644,02	24,86	0,00	0,09
		32	327,98	19,31	327,98	19,22	0,00	0,09
		37	174,94	14,11	174,94	14,16	0,00	-0,05
		42	91,32	9,26	91,32	9,29	0,00	-0,03
		47	48,11	6,19	48,11	6,26	0,00	-0,07
	BQSquare_ 416x240_60_val	22	1326,46	34,71	1326,46	34,77	0,00	-0,06
		27	507,22	32,58	507,22	32,56	0,00	0,01
		32	226,60	29,75	226,60	29,74	0,00	0,01
		37	117,35	24,08	117,35	24,08	0,00	0,00
		42	65,23	14,97	65,23	14,96	0,00	0,01
		47	36,20	9,09	36,20	9,09	0,00	0,00
	BlowingBubbles_ 416x240_50_val	22	1601,26	67,61	1473,64	67,04	127,62	0,57
		27	799,48	65,14	735,62	65,97	63,86	-0,83
		32	386,26	64,97	355,86	64,21	30,40	0,76
		37	187,76	52,73	174,37	53,98	13,39	-1,25

		42	89,29	28,06	84,24	31,87	5,05	-3,81
		47	41,04	27,65	39,92	29,89	1,12	-2,25
	RaceHorses_ 416x240_30_val	22	876,56	42,93	868,26	43,50	8,30	-0,57
		27	454,85	41,57	454,32	42,35	0,53	-0,78
		32	235,41	38,56	235,23	38,50	0,19	0,07
		37	126,33	33,65	126,39	33,81	-0,06	-0,16
		42	67,51	23,82	68,39	25,22	-0,89	-1,40
		47	37,16	13,05	37,27	12,11	-0,11	0,94
ClassE 100%	FourPeople_ 1280x720_60_val	22	1500,83	26,71	1474,69	26,58	26,14	0,13
		27	832,55	25,46	823,68	25,93	8,86	-0,47
		32	480,15	26,52	475,10	25,90	5,05	0,63
		37	284,16	24,28	279,90	23,85	4,25	0,43
		42	168,82	24,14	166,01	23,26	2,81	0,89
		47	96,73	20,70	96,59	20,93	0,13	-0,23
	Johnny_ 1280x720_60_val	22	934,49	61,58	876,74	61,31	57,75	0,27
		27	477,66	61,70	454,32	60,48	23,34	1,22
		32	262,91	61,45	252,64	61,30	10,28	0,15
		37	154,26	63,26	149,71	62,17	4,55	1,09
		42	94,03	60,13	90,70	61,05	3,33	-0,91
		47	58,74	53,74	57,92	54,26	0,81	-0,51
	KristenAndSara_ 1280x720_60_val	22	1254,01	26,09	1254,01	26,17	0,00	-0,07
		27	626,27	27,17	626,27	27,14	0,00	0,03
		32	350,81	26,59	350,81	26,65	0,00	-0,06
		37	209,48	24,55	209,48	24,51	0,00	0,04
		42	128,35	22,42	128,35	22,41	0,00	0,02
		47	79,07	22,88	79,07	22,82	0,00	0,06

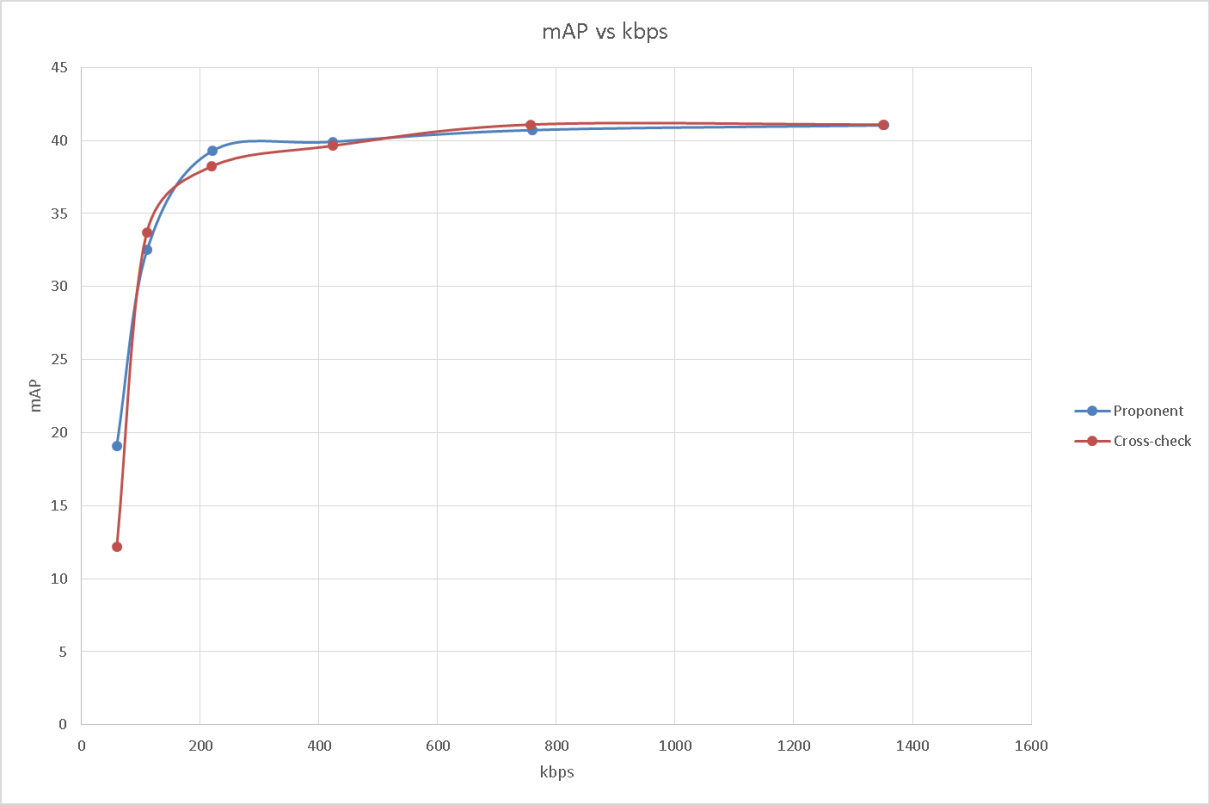

Fig. 5. Object detection result on SFU Traffic_2560x1600_30_val sequence

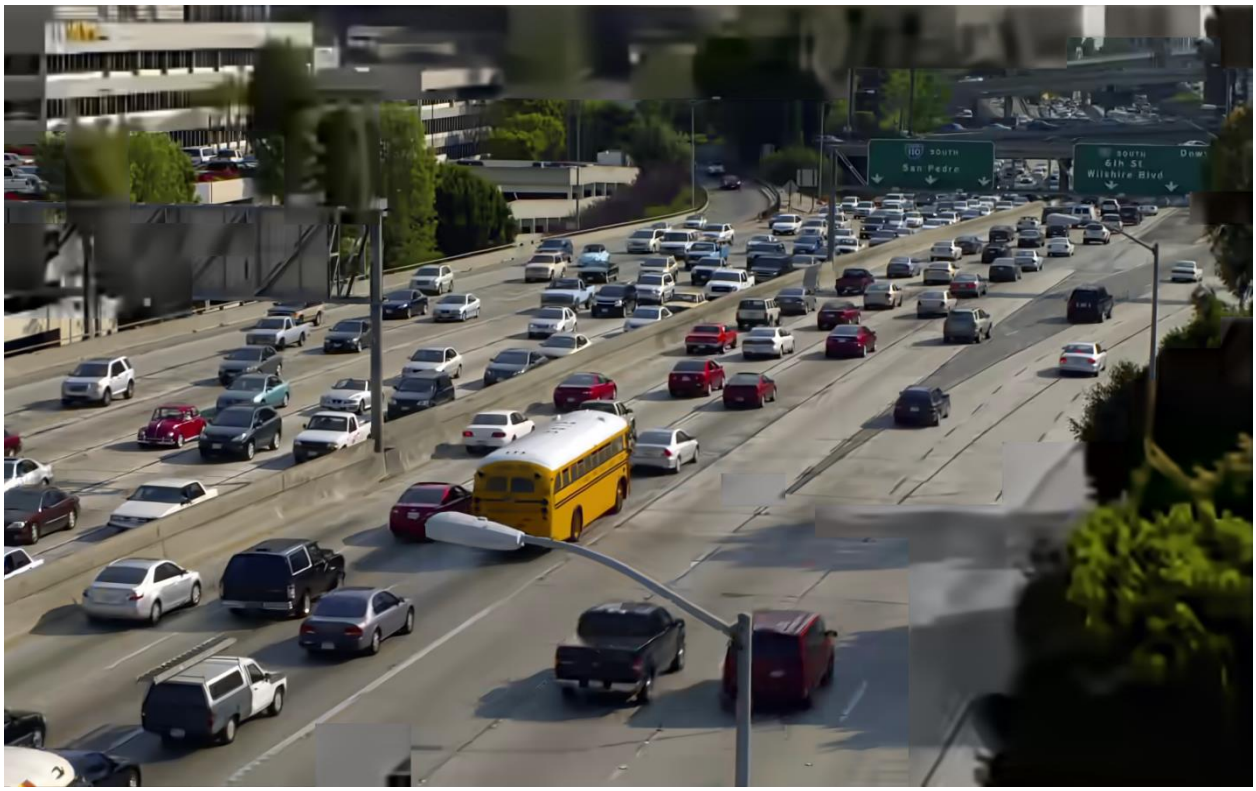

Fig 6. Exemplary decoded frame from Traffic_2560x1600_30_val sequence (frame 021, qp 47)

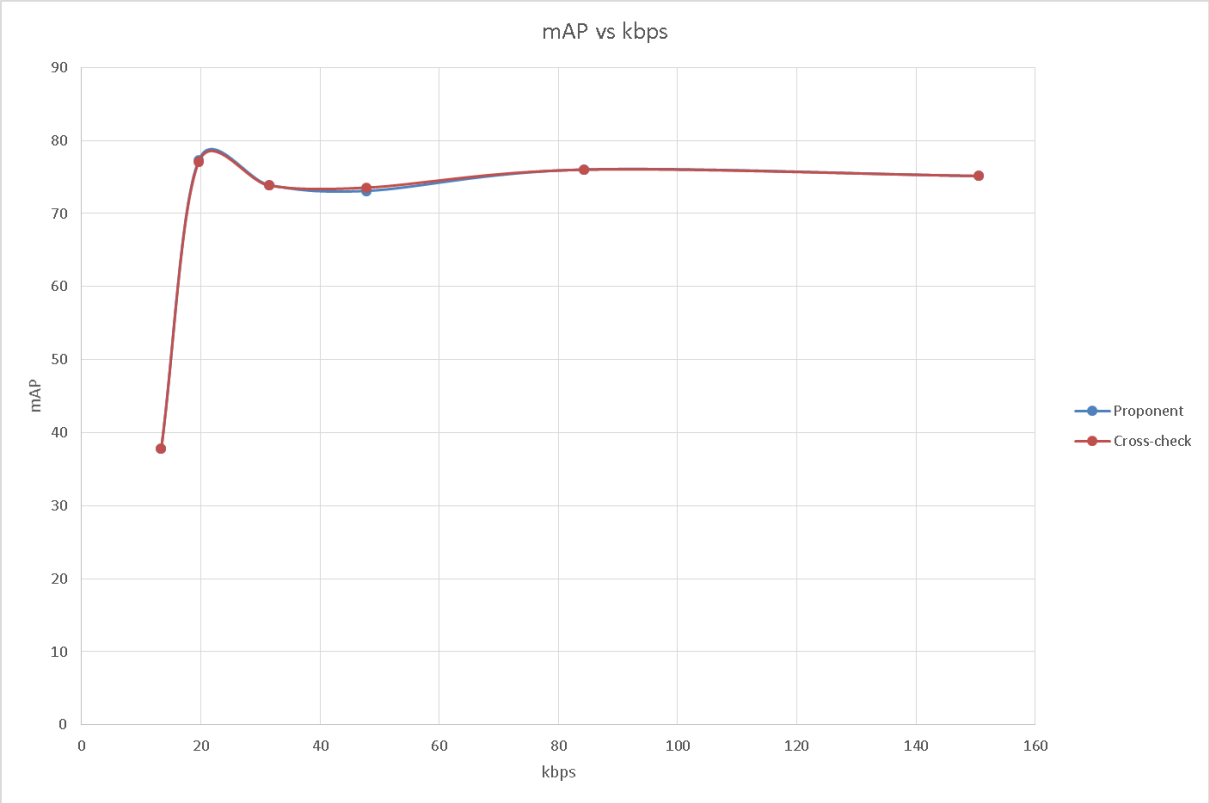

Fig. 7. Object detection result on SFU Kimono_1920x1080_24_val sequence

Fig. 8. Exemplary decoded frame from Kimono_1920x1080_24_val sequence (frame 004, qp 42)

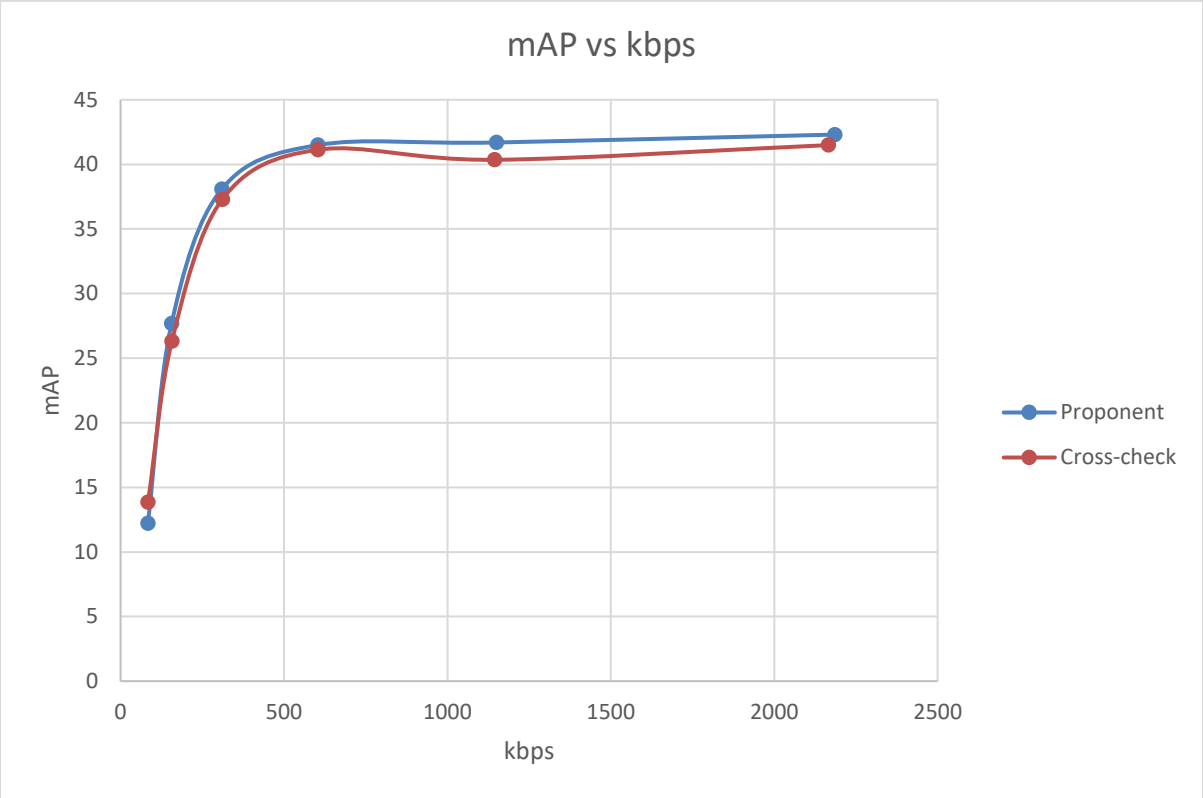

Fig. 9. Object detection result on SFU BasketballDrive_1920x1080_50_val sequence

Fig. 10. Exemplary decoded frame from BasketballDrive_1920x1080_50_val sequence (frame 017, qp 42)

Table. 3. Object detection results on SFU datasets (divided into sequence classes)

Scale			CE.1.3.		OUR Cross-check		Difference	
	Class	RP	kbps	mAP	kbps	mAP	kbps	mAP
100%	Class AB	0	1392,17	54,56	1396,132	42,88	-3,96	11,68
		1	731,19	53,02	733,561	42,38	-2,37	10,64
		2	385,35	51,97	386,783	41,12	-1,43	10,85
		3	199,82	49,70	200,613	37,29	-0,79	12,41
		4	99,67	42,34	100,151	26,71	-0,48	15,63
		5	52,30	20,39	52,367	8,78	-0,07	11,61
	Class C	0	1577,53	45,94	1577,343	41,86	0,18	4,08
		1	793,94	43,43	793,731	40,15	0,21	3,28
		2	417,23	39,45	416,813	34,38	0,41	5,07
		3	218,30	34,59	218,064	29,36	0,23	5,24
		4	110,93	22,82	110,932	14,84	0,00	7,98
		5	55,11	14,70	55,108	8,14	0,00	6,57
	Class D	0	1265,14	43,17	1014,094	37,88	251,05	5,29
		1	601,39	41,06	471,715	36,01	129,68	5,05
		2	294,06	38,15	227,611	32,80	66,45	5,35
		3	151,59	31,14	116,665	26,50	34,93	4,65
		4	78,34	19,03	60,198	14,44	18,14	4,59
		5	40,63	14,00	31,057	9,17	9,57	4,82
	Class E	0	1229,77	38,13	1201,812	31,12	27,96	7,01
		1	645,49	38,11	634,757	31,57	10,73	6,54
		2	364,62	38,19	359,515	32,80	5,11	5,39
		3	215,97	37,36	213,031	31,94	2,93	5,42
		4	130,40	35,57	128,351	32,44	2,05	3,12
		5	78,18	32,44	77,861	30,01	0,32	2,43

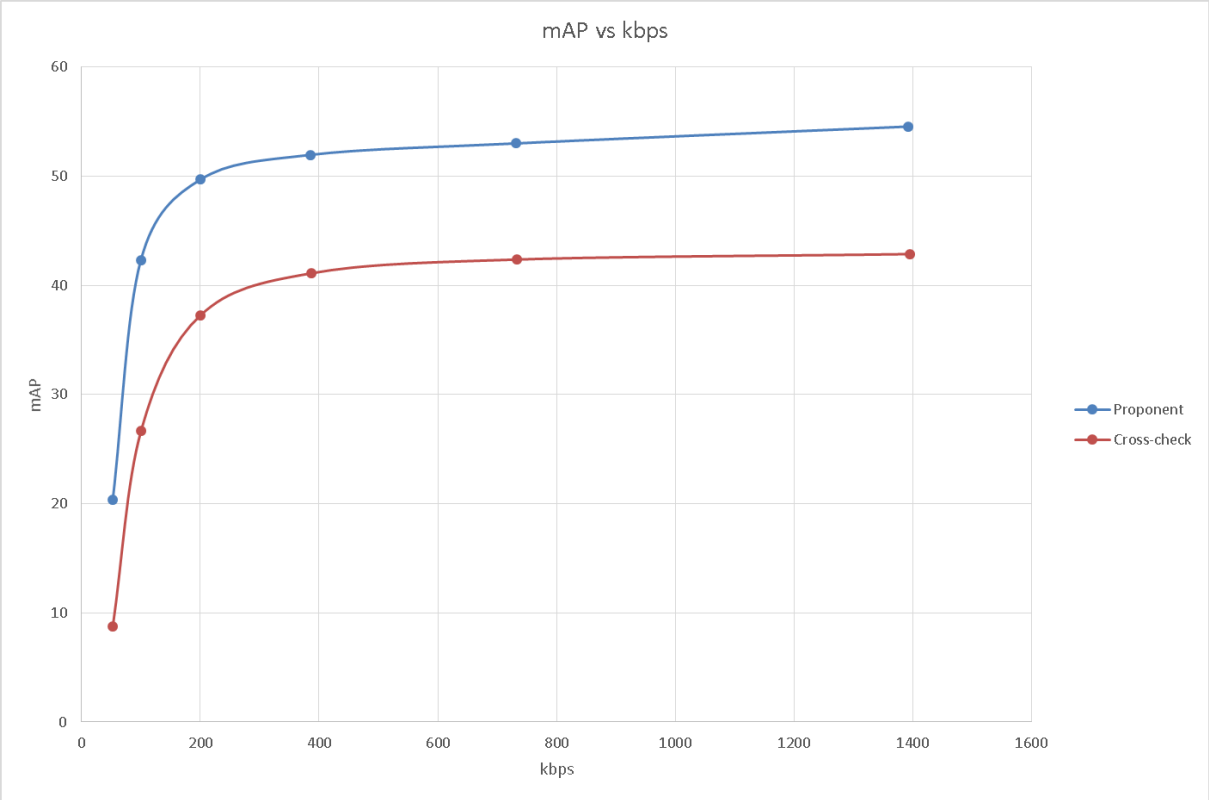

Fig. 11. Object detection result on SFU sequences Class AB

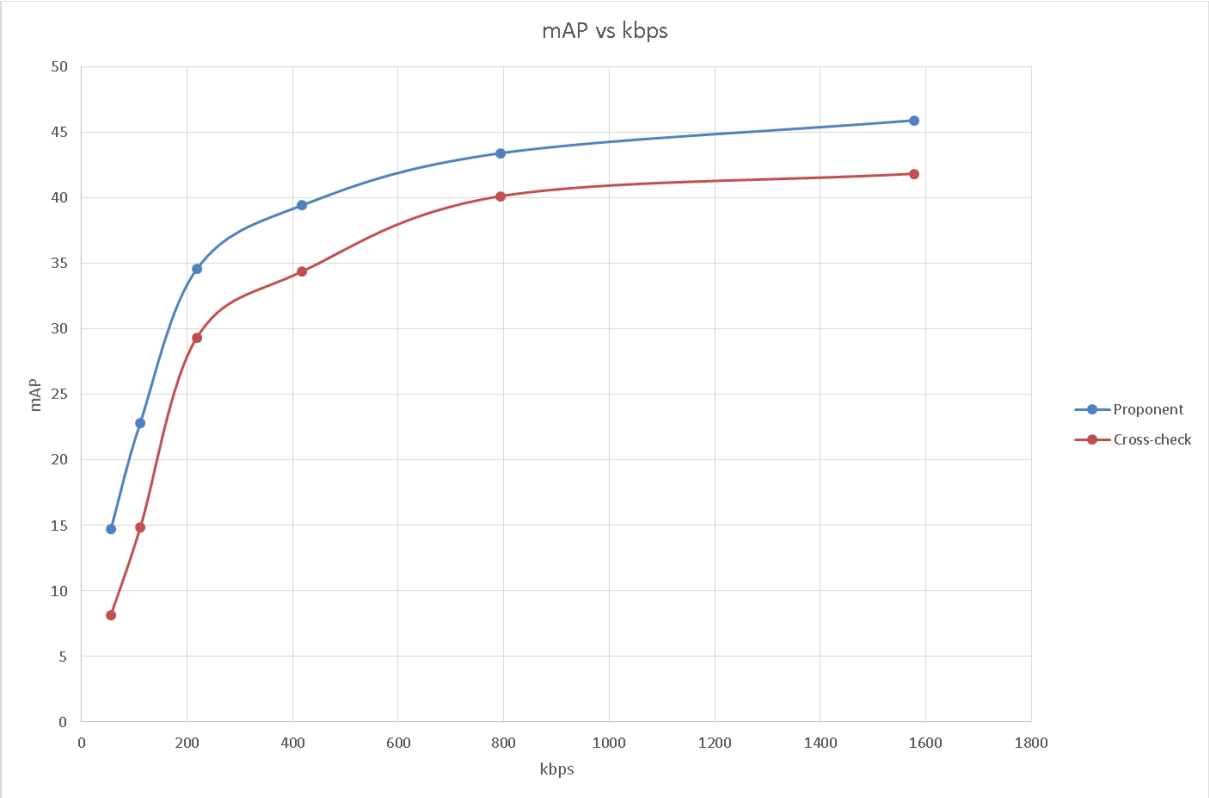

Fig. 12. Object detection result on SFU sequences Class C

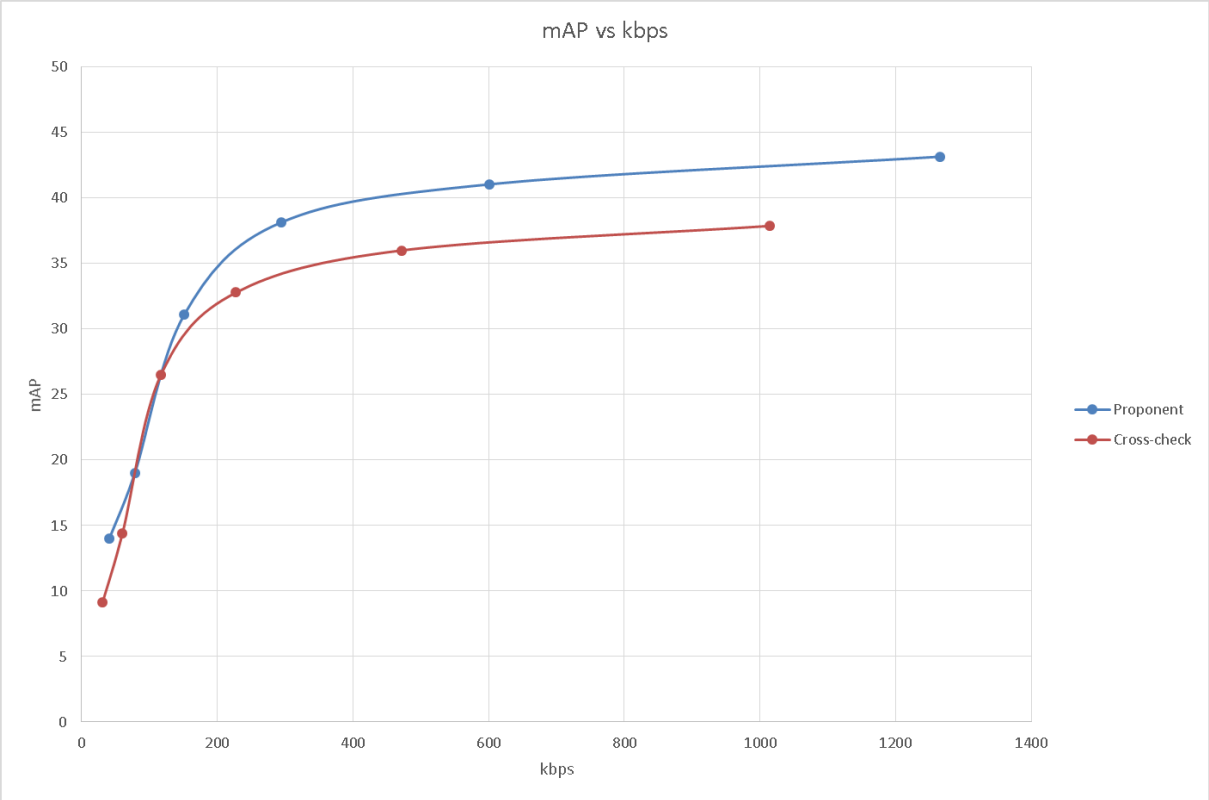

Fig. 13. Object detection result on SFU sequences Class D

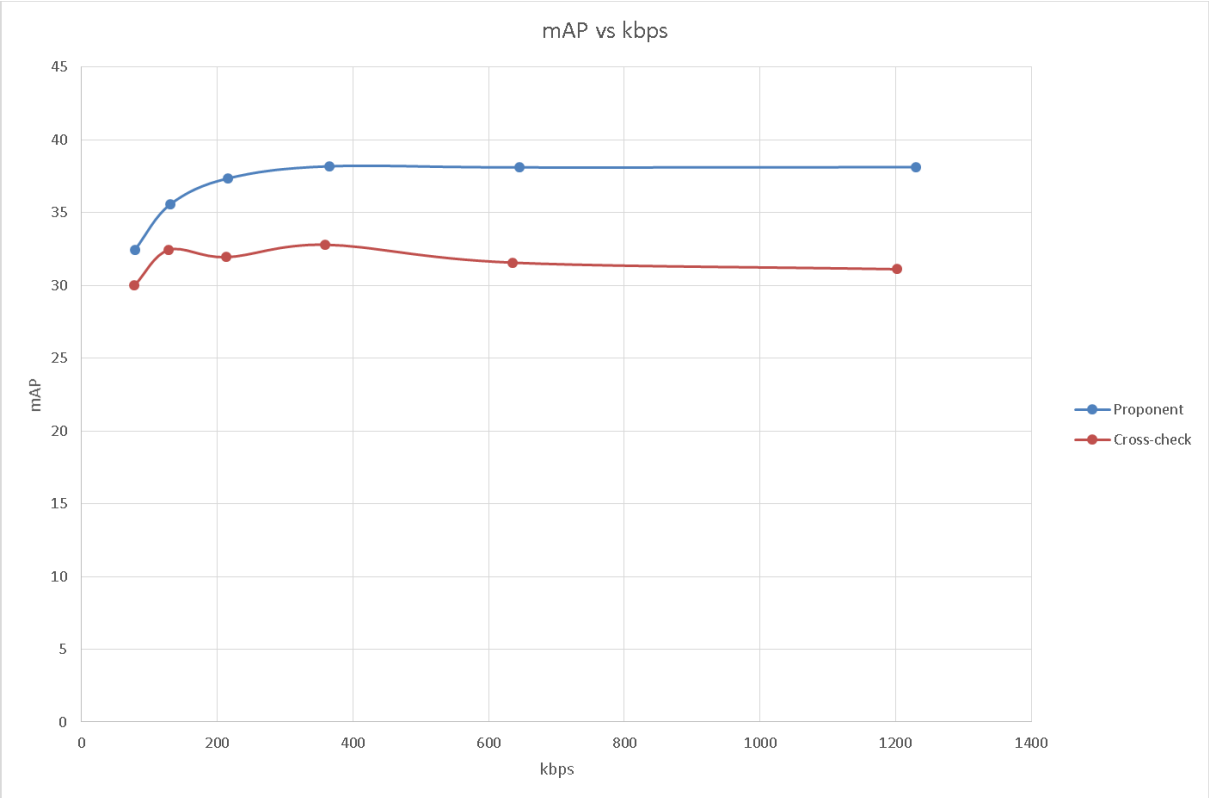

Fig. 14. Object detection result on SFU sequences Class E

2.2. Instance Segmentation

Table. 4. Instance segmentation results on TVD dataset

			CE1.3.		OUR cross-check		Difference	
Scale	Dataset	QP	BPP	mAP	BPP	mAP	BPP	mAP
100%	TVD	22	0,153	43,001	0,154	42,574	0,000	0,428
		27	0,090	41,567	0,090	41,451	0,000	0,117
		32	0,052	37,201	0,052	37,411	0,000	-0,209
		37	0,029	35,075	0,029	34,812	0,000	0,264
		42	0,016	24,874	0,016	24,800	0,000	0,074
		47	0,010	17,075	0,010	17,088	0,000	-0,013

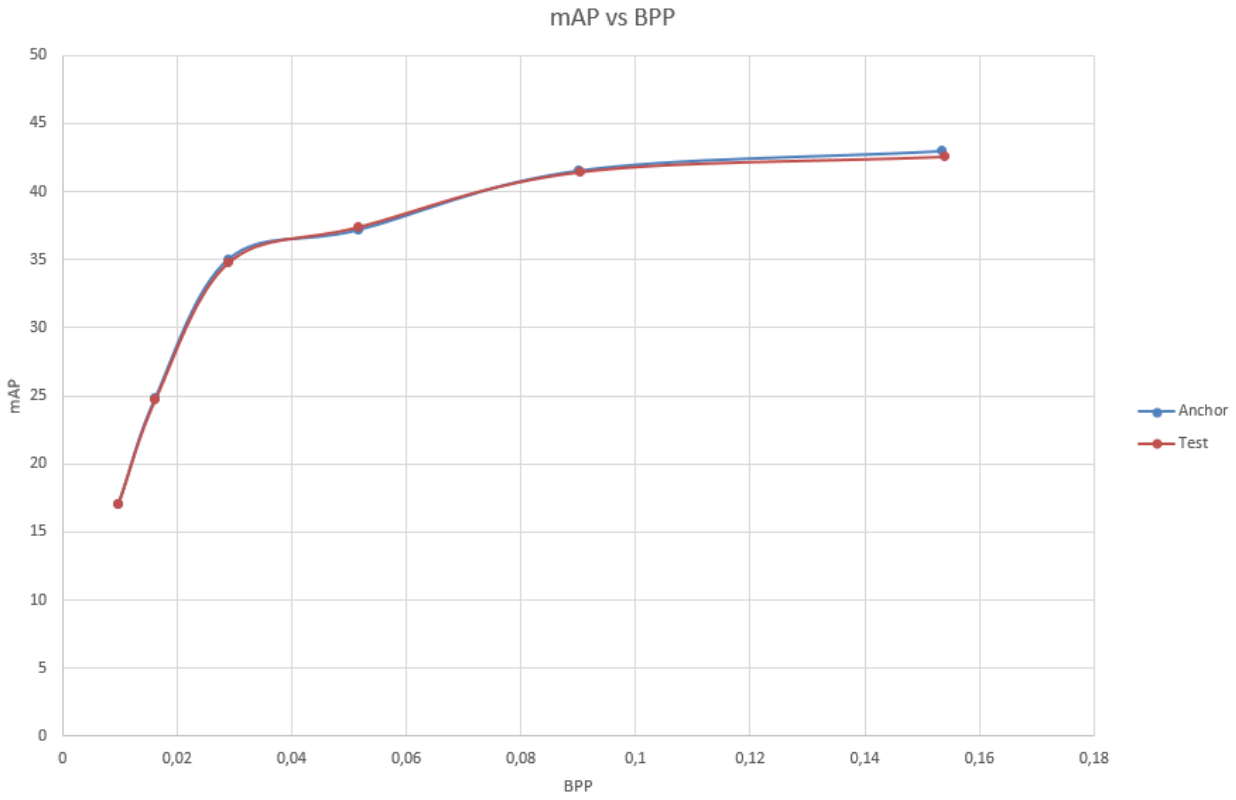

Fig. 15. Object segmentation result on TVD dataset

3. Issues

During our crosscheck, we have come across the following concerning issues:

- No specified hardware requirements or differences in hardware requirements between proponent's framework and default reference software [2].
- Path issues in scripts provided by Proponent, we had to copy files from *CE1.3/CE1.3* directory to the main folder so it could work properly. Example error:

  No such file or directory: './vcmrs/InnerCodec/VTM_CU_based/bin/EncoderAppStatic'.

  It occurred when we tried running scripts from *CE1.3/CE1.3* folder, where they were by default.

- Different default hardware configuration per encoder/decoder configuration file for each dataset. Some of them required 2 GPUs, some even 6 GPUs so we had to adjust it for a single GPU setup.
- Encoding speed scales poorly with increasing encoding quality. For example, according to information we were given by Proponent, encoding images from TVD dataset with QP 22 can take even 2 hours per image while using workstation equipped with AMD Threadripper or Intel i9 CPU and Nvidia Quadro GPU. Perhaps it is not as bad while encoding many QPs at once but our machines were not good enough to allow it. On the other hand, decoding times for images datasets were acceptable. For SFU sequences encoding times were quite high. In many cases, encoding lower QPs often took more than 6 hours per QP. As mentioned before, it may not be as bad when encoding multiple QPs at once but in our case it was not available. Decoding times were faster but many sequences still needed around 25 minutes per QP to be decoded.
- Encoding scripts had some variables which could improve encoding speed, for example `processes_per_gpu` in *Scripts/encode_flir.py*. Perhaps we could improve our encoding speed by properly reconfiguring them but since there was no description for them, we did not have spare time to check every configuration for best outcome.
- Temporary files were taking way too much hard drive space. For example, encoding ~2200 images from OpenImages dataset while using QP 47 required over 110 GB of hard drive space. By default, they are stored in */tmp* catalog, but when hard drive runs out of free space it starts to create temporary files in encoder catalog. We are not sure if is general problem of using reference software framework or adding too many extra steps/plugins.
- For some reason script FLIR.sh required addition copy of *model_final.pth* in input data folder.
- Script TVD-video does not work by default.

  First encountered error:

  File "CE1.3/vcmrs/ROI/MachineAttention/GOP_attention.py", line 38, in read_multi_json_files

  return json_data, total_json_info

  UnboundLocalError: local variable 'json_data' referenced before assignment.

After fixing error above we encountered another error:

File "CE1.3/vcmrs/ROI/MachineAttention/GOP_attention.py", line 122, in calc_GOP_attention_region

total_objectness=np.zeros((json_data['height'], json_data['width']))

TypeError: list indices must be integers or slices, not str

We have not tried to fix this part and we do not know if there are any further issues.

- By default, decoding in script *SFU.sh* was not working - we could run this script without visible errors but we were not getting any decoded images or information in log files. We had to modify it before it started working properly.
- BPPs calculated for FLIR and TVD-images datasets match bitrates provided to us. Bitrates calculated for SFU dataset are slightly different than those provided to us with the biggest difference for BlowingBubbles sequence and in result for class D.
- Our evaluation mAP values for FLIR, TVD-images and SFU datasets are more or less different than those provided to us.
- The proposed mAP results for the SFU-HW classes were determined on the basis of the arithmetic average of the mAP results for individual sequences. This approach is not in accordance with the assumptions of the provided software, which causes large discrepancies in our cross-check.

4. Summary:

In general, most of results provided by the proponent were more or less similar to our results with some exceptions described above. The only identical results were BPPs for FLIR and TVD-images datasets.

References

[1] Yegi Lee, Shin Kim, Kyoungro Yoon(Konkuk Univ.), Hanshin Lim, Sangwoon Kwak, Hyon-Gon Choo, Soon-heung Jung, Jeogil Seo(ETRI), “[VCM] CE1.3-Machine Attention-based Coding”, January 2022, Online

[2] Honglei Zhang (Nokia), “[VCM] Introduction to the VCM reference software (VCM-RS)”, January 2023, Online