Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
32nd Meeting, Hannover, DE, 13-20 October 2023

Document: JVET-AF0023
MPEG: m65062

Title: **EE1: Summary report of exploration experiment on neural network-based video coding**

Status: Input document to JVET

Purpose: Proposal

Author(s) or Contact(s):
Elena Alshina | elena.alshina@huawei.com
Franck Galpin | franck.galpin@interdigital.com
Yun Li | yli30@qti.qualcomm.com
Dmytro Rusanovskyy | dmytro.rusanovskyy@nokia.com
Maria Santamaria | maria.santamaria_gomez@nokia.com
J. Ström | jacob.strom@ericsson.com
Liqiang Wang | liqiangwang@tencent.com
Zhihuang Xie | xiezhihuang@oppo.com
Jens-Rainer Ohm | ohm@ient.rwth-aachen.de | submitter

Source: E. Alshina, F. Galpin, Y. Li, D. Rusanovskyy, M. Santamaria, J. Ström, L. Wang, Z. Xie

---

HEADER EXCERPT FROM SOURCE DOCUMENT

Abstract

This report summarizes the activities of the Exploration Experiment 1 (EE1) performed between the JVET-AE and JVET-AF meetings to evaluate Neural Network-based Video Coding (NNVC) technologies, analyze their performance, evaluate their complexity aspects, and clarify training procedure. Two teleconferences have been conducted during this meeting circle.

Introduction

This round of EE1 tests included:

EE1-0.x: Unified LOP (based on JVET-AE0281)

EE1-1.1.x category: architectural changes for HOP (based on JVET-AE0160 , JVET-AE0164).

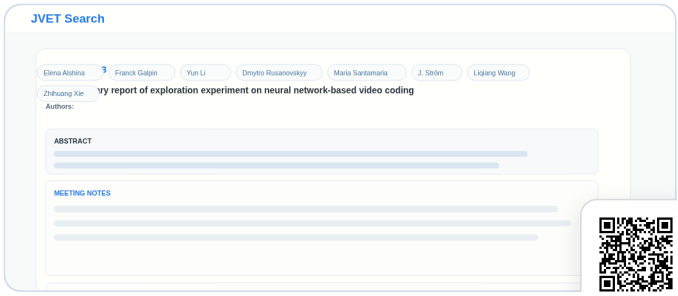
Built from JVET-AF0023-v1.docx

LINKS

CARD

JVET-AF0023

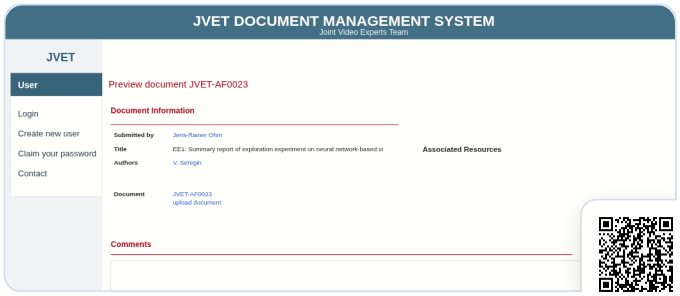
http://vexpertise.stbit.ru/documents/JVET-AF0023



SOURCE

external source

https://jvet-experts.org/doc_end_user/current_document.php?id=13469



ARCHIVE

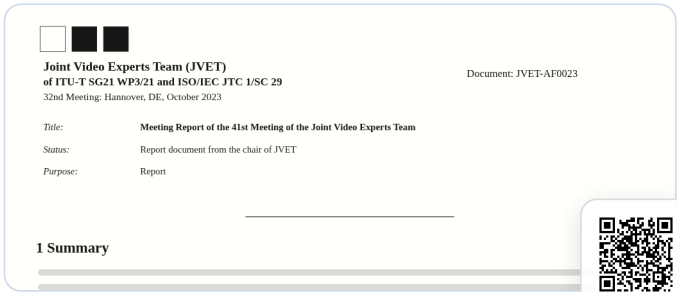
archive

https://jvet-experts.org/doc_end_user/documents/32_Hannover/wg11/JVET-AF0023-v3.zip

MEETING NOTES

external meeting notes

http://vexpertise.stbit.ru/meeting-notes/source/meeting_32_af



MEETING NOTES

JVET-AF0023 EE1: Summary report of exploration experiment on neural network-based video coding [E. Alshina, F. Galpin, Y. Li, D. Rusanovskyy, M. Santamaria, J. Ström, L. Wang, Z. Xie]

This report summarizes the activities of the Exploration Experiment 1 (EE1) performed between the JVET-AE and JVET-AF meetings to evaluate Neural Network-based Video Coding (NNVC) technologies, analyze their performance, evaluate their complexity aspects, and clarify training procedure. Two teleconferences have been conducted during this meeting circle.

This round of EE1 tests included:

- EE1-0.x: Unified LOP (based on JVET-AE0281)
- EE1-1.1.x category: architectural changes for HOP (based on JVET-AE0160 , JVET-AE0164).
- EE1-1.2.x category: filter usage aspects (based on JVET-AE0072, JVET-AE0093, JVET-AE0161, JVET-AE0238).
- EE1-2.x NN-Inter (based on JVET-AE0112 investigating complexity reduction possibilities).
- EE1-3.x NN-Intra (based on JVET-AE0144, awaiting training cross-check).

All tests in EE1 were advised to use NNVC-6.0 as code base (unless it is not possible). The anchor for EE1 test is the default configuration of NNVC-6.0 as defined by AhG11/AhG14 (NN-intra and low complexity NN-filter enabled by default) in JVET-AE2016. Anchor performance and reference point for HOP NN-filters will be provided by AhG14.

Tests targeting to LOP replacement in NNVC (EE1-0) should report results vs NNVC-6.0 default configuration (LOP-1 filter and NN-Intra are enabled by default). Tests targeting improvement of HOP (EE1-1 category) should report results in comparison to NNVC-6.0 with HOP-1 (instead LOP-1) and NN-Intra enabled. HOP-1 architecture and reproducible training procedure description can be found in JVET-AE2019 and readme.md.

NNVC has two NN-filters, so can be configures for Low Operation Point (LOP) and High Operation Point (HOP), in both NN-Intra is enabled. Performance of those two configurations relatively to VTM is summarized in the table below. Run time is not always very reliable and accurate, but can give group extra information about complexity, so run time was added to the tables.

Two configuration of NNVC tested vs VTM, NN-Intra enabled in both.

Test	Doc#	# Par M	kMAC /pxl	Random Access cfg.					All Intra cfg.				
				Y	U	V	Enc	Dec	Y	U	V	Enc	Dec
LOP	JVET-AE0014	0.2+1.3	17+8	**-6.6%**	-7.7%	-7.3%	1.3	76	**-7.8%**	-8.6%	-8.9%	2.0	45
HOP	JVET-AE0014	1.5+1.3	477+8	**-11.6%**	-25.7%	-26.0%	1.8	476	**-10.8%**	-22.2%	-23.4%	2.1	269

By default in NNVC-6.0 NN-Intra and Low Operation Point (LOP) NN-filter are enabled. Proposals tested against NNVC-6.0 are listed in the next table below.

Proposals tested against NNVC default configuration (LOP and NN-Intra enabled)

Test	Doc#	# Par M	kMAC /pxl	Random Access cfg.					All Intra cfg.				
				Y	U	V	Enc.	Dec.	Y	U	V	Enc.	Dec.

NNVC-6-LOP	JVET-AE0014	0.2	17	**0.0%**	0.0%	0.0%	100%	100%	**0.0%**	0.0%	0.0%	100%	100%
Targeting replacement of LOP filter in NNVC													
EE1-0	JVET-AF0043	0.05	17	**-0.3%**	-4.3%	-4.3%	96%	114%	**-0.2%**	-4.6%	-4.6%	97%	119%
Content adaptation with overfitting at encoder side													
EE1-1.2.2	JVET-AF0056	0.2	17	**-1.2%**	-6.4%	-5.5%	99%	105%					
NN-Inter (performance on top of LOP & NN-Intra)													
EE1-2.1	JVET-AF0208	3.8	504	**-2.3%**	-4.2%	-3.8%	133%	1204%					
Targeting replacement of NN-Intra filter in NNVC													
EE1-3.1	JVET-AF0139	1.3	4.8	**0.2%**	0.1%	0.2%	99%	99%	**0.5%**	0.6%	0.7%	95%	94%

The majority of EE1 tests are focusing on improvement of High Operation Point (HOP) complexity-performance trade-off. The natural anchor for those tests is NNVC-6.0 with HOP and NN-Intra enabled. Test results relatively to this anchor are summarized in the next table below.

Proposals tested against NNVC- high configuration (HOP and NN-Intra enabled)

Test	Doc#	#Par M	kMAC /pxl	Random Access cfg.					All Intra cfg.				
				Y	U	V	Enc.	Dec.	Y	U	V	Enc.	Dec.
NNVC-6-HOP	JVET-AF0014	1.45	477	**0.0%**	0.0%	0.0%	100%	100%	**0.0%**	0.0%	0.0%	100%	100%
Group convolutions in BBB (Back Bone Block)													
EE1-1.1.1-t1	JVET-AF0153	1.28	422	**0.3%**	2.0%	1.5%			**0.1%**	1.1%	1.0%		
EE1-1.1.4b1	JVET-AF0182	1.23	406	**0.3%**	2.7%	1.0%	120%	155%	**0.1%**	0.7%	0.7%	111%	157%
EE1-1.1.4b2	JVET-AF0182	1.27	419	**0.3%**	1.1%	1.2%	124%	172%	**0.2%**	0.9%	0.9%	112%	175%

EE1-1.1.4b3	JVET-AF0182	1.05	349	**0.6%**	2.6%	2.4%	135%	204%	**0.3%**	1.3%	1.1%	118%	210%
Number of channels reduction in BBB (Back Bone Block)													
EE1-1.1.2b	JVET-AF0102	1.3	426	**0.1%**	1.0%	0.4%	82%	71%	**0.1%**	0.2%	0.4%	83%	66%
EE1-1.1.1-t2	JVET-AF0153	1.35	442	**0.3%**	1.3%	0.9%			**0.2%**	0.8%	1.1%		
EE1-1.1.5	JVET-AF0192	1.2	392	**0.4%**	1.7%	1.2%	72%	56%	**0.2%**	1.0%	0.8%	77%	53%
“head” block simplifications													
EE1-1.1.4a	JVET-AF0181	1.45	457	**0.2%**	0.9%	0.5%	95%	86%	**0.1%**	0.2%	0.4%	98%	89%
EE1-1.1.3	JVET-AF0103	1.45	472	**0.0%**	1.0%	0.3%	98%	96%	**0.1%**	0.3%	0.3%	95%	92%
Separation of NN-filters for Luma and Chroma													
EE1-1.1.4c	JVET-AF0183	1.42	473	**-0.3%**	1.5%	1.9%	129%	202%	**-0.4%**	0.3%	0.5%	112%	206%
EE1-1.1.4d	JVET-AF0183	1.23	390	**0.0%**	1.8%	2.0%	137%	221%	**-0.2%**	-0.1%	0.2%	116%	232%
Filter usage aspects (also applicable to LOP)													
EE1-1.2.1	JVET-AF0085	1.45	477	**0.0%**	-0.2%	-0.5%	100%	101%	**0.0%**	-0.1%	0.0%	100%	99%
EE1-1.2.3	JVET-AF0154	1.45	477	**-0.1%**	-0.6%	-0.8%	97%	99%					
EE1-1.2.4	JVET-AF0205	1.45	477	**0.0%**	-0.1%	-0.1%	100%	99%	**0.0%**	-0.1%	-0.1%	100%	100%
EE1-1.2.6	JVET-AF0086	1.45	477	**-0.1%**	-0.5%	-0.7%	98%	99%					

Some proponents managed to complete LDB test, results are summarized in the next table below.

Proposals tested against NNVC- high configuration (HOP and NN-Intra are enabled) in LDB configuration.

Test	Doc#	#Par M	kMAC /pxl	Low Delay-B cfg.				
				Y	U	V	Enc.	Dec.
NNVC-6-HOP	JVET-AF0014	1.45	477	**0.0%**	0.0%	0.0%	100%	100%
EE1-1.1.1-t1	JVET-AF0153	1.28	422	**0.2%**	-0.1%	-6.2%		
EE1-1.1.4b1	JVET-AF0182	1.23	406	**0.2%**	0.0%	-5.8%	121%	157%
EE1-1.1.4b2	JVET-AF0182	1.27	419	**0.2%**	-0.4%	-7.4%	127%	174%
EE1-1.1.1-t2	JVET-AF0153	1.35	442	**0.1%**	-0.5%	-8.2%		
EE1-1.1.5	JVET-AF0192	1.2	392	**0.3%**	-0.8%	-7.2%		
EE1-1.1.4a	JVET-AF0181	1.45	457	0.3%	-1.6%	-8.4%	94%	85%
EE1-1.2.1	JVET-AF0085	1.45	477	**-0.3%**	-1.8%	-4.9%	100%	105%
EE1-1.2.3	JVET-AF0154	1.45	477	**-0.4%**	-2.6%	-2.2%	98%	100%
EE1-1.2.6	JVET-AF0086	1.45	477	**-0.4%**	-2.1%	-1.4%	98%	101%

Observations of those test results are as follows:

- Reduction of kMAC/pxl not necessarily leads to reduction of encoding/decoding run-time (often causes opposite effect).
- Strangely almost all proposals tested in LDB cfg demonstrated significant gain in PSNR-V. Likely this indicates some problem with NNVC-HOP training or usage.

Observations on tests:

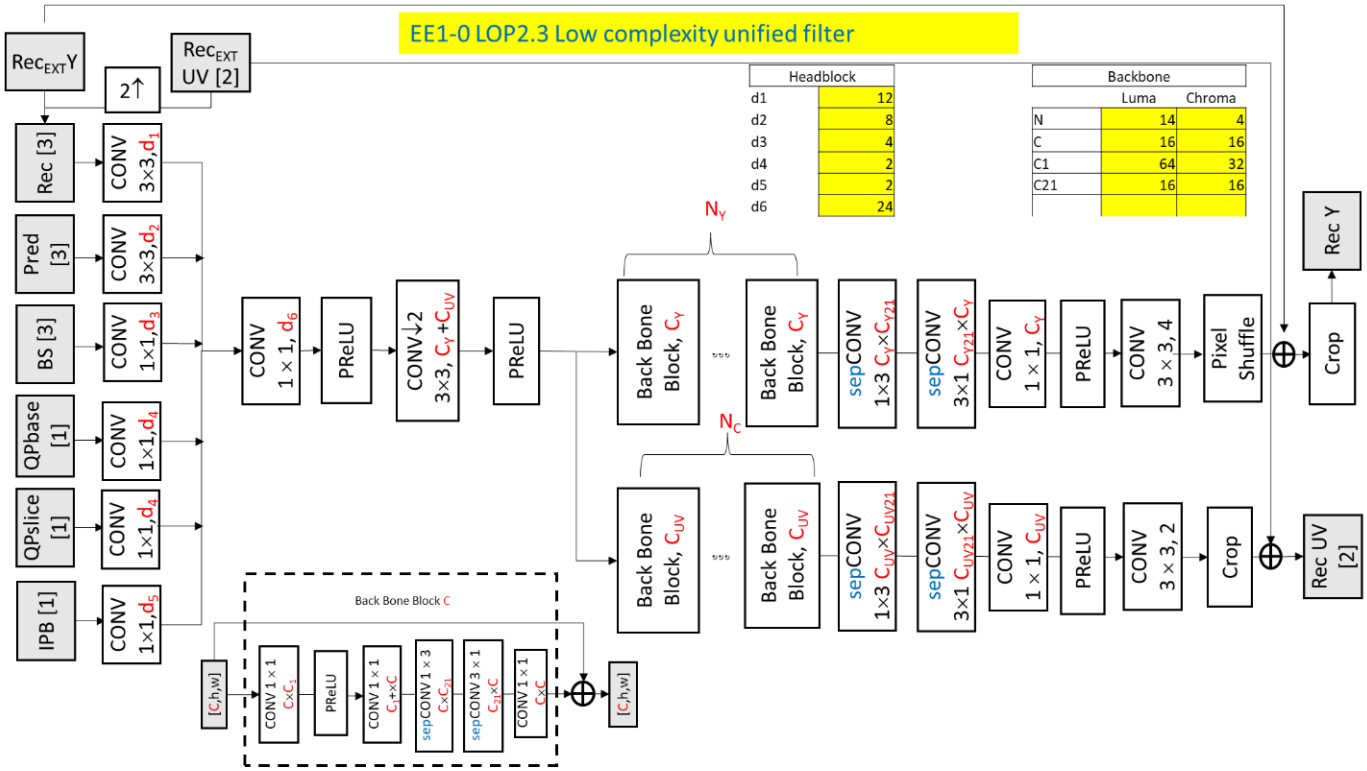
- Majority of proposed HOP NN-filter simplifications do not contradict to each other, potentially can be used together (the only exception is group convolution size, group should choose between group size 2 or 4).
- Among filter usage aspects (flipping, rotating, residual adjustment…) noticeable (0.3-0.4%) gain is observed in LDB cfg.

Mode details about EE1 tests can be found in presentation attached to this contribution.

Tests description and results

EE1-0 tests 0.1 to 0.4

Input contribution: JVET-AF0043 **AhG11/EE1: Status of the joint EE1-0 (LOP.2) training** D. Rusanovskyy, Y. Li (Qualcomm), T. Shao, P. Yin (Dolby), J. N. Shingala, A. Shyam, A. Suneja, S. P. Badya (Ittiam), J. Li, Y. Li, C. Lin, K. Zhang, L. Zhang (Bytedance), R. Chang, L. Wang, X. Xu, S. Liu (Tencent)



Proposed unified architecture for LOP EE1 test

Training and inference were cross-checked since training was conducting by multiple parties: Qualcomm, Ittiam, Dolby (no separate documents were deemed needed).

After Stage 3 training, selected LOP 2.3 candidate model demonstrated BD-rate change of: {-4.7%, -9.8, -10.1%} and {-5.3%, -10.9%, -10.4%} vs VTM for AI and RA configurations, respectively.

Comparing to NNVC anchor (NN-Intra ON), LOP2.3 with enabled NN-Intra demonstrated BD-rate change of {-0.2%, -4.6%, -4.6%} and {-0.3%, -4.3%, -4.3%} for AI and RA, respectively. In this test, filter LOP2.3 utilizes a single model with 0.05M parameters, whereas the Anchor uses 4 models, total size of 0.2M.

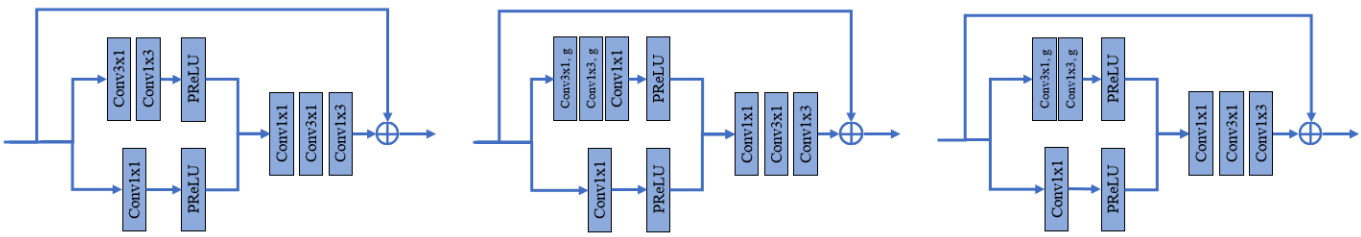
Additional BD-rate gain of {-0.0%, -0.5%, -0.2%} and {-0.0%, -1.3%, -0.3%} for AI and RA, respectively, is reported in the sub-test EE1-0.4, targeting improved filter usage (interface). Aspects of common interest:

- finding in training strategy,
- layer quantization

Recommendation: to be presented in details as strong candidate for adoption and source of useful information.

EE1-1.1.1

Input contribution: JVET-AF0153 EE1-1.1.1: Optimization for complexity-performance trade-off of HOP network R. Chang, L. Wang, X. Xu, S. Liu (Tencent)



Filter architecture: HOP-based

Training strategy: HOP

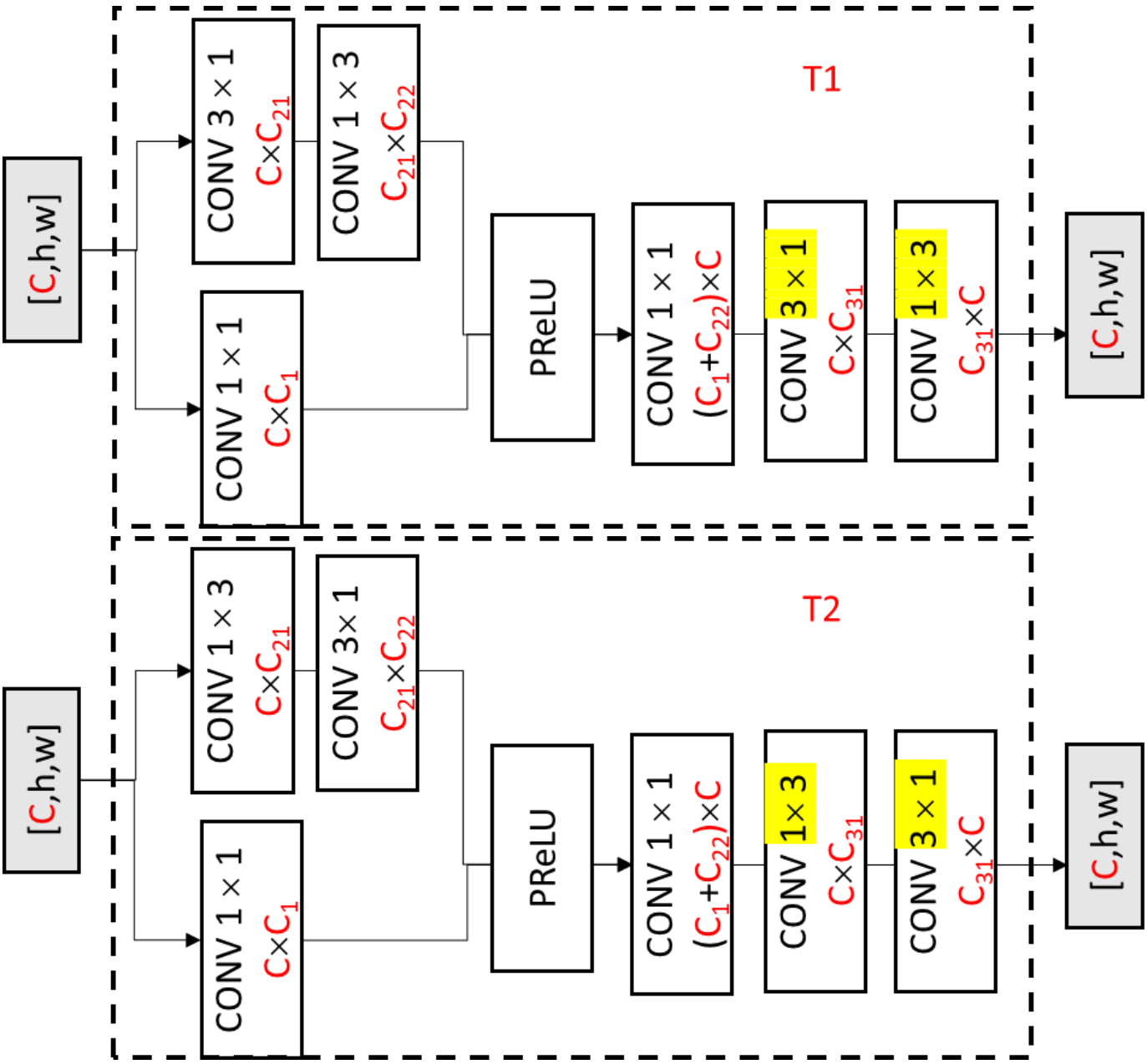
Training conducted: Tencent

Cross-check: Inference and training cross-check of test 1 are conducted, cross-checker: OPPO (JVET-AF0266).

Notes: This test includes subtests targeting to optimize the complexity-performance trade-off of residual block structures and network architecture. The improved residual blocks (HOP residual block with depth-wise separable convolution and HOP residual block with group convolution) and split architecture from JVET-AE0160 are to be tested. This test may include the subtests by combining the elements from JVET-AE0160 and the subtests planned in test EE1-1.5 from JVET-AD2023.

EE1-1.1.2

Input contribution: JVET-AF0102 EE1-1.1.2 Complexity-performance tradeoff of decomposition
D. Rusanovskyy, Y. Li, M. Karczewicz (Qualcomm)



Filter architecture: HOP-based.

Training strategy: HOP

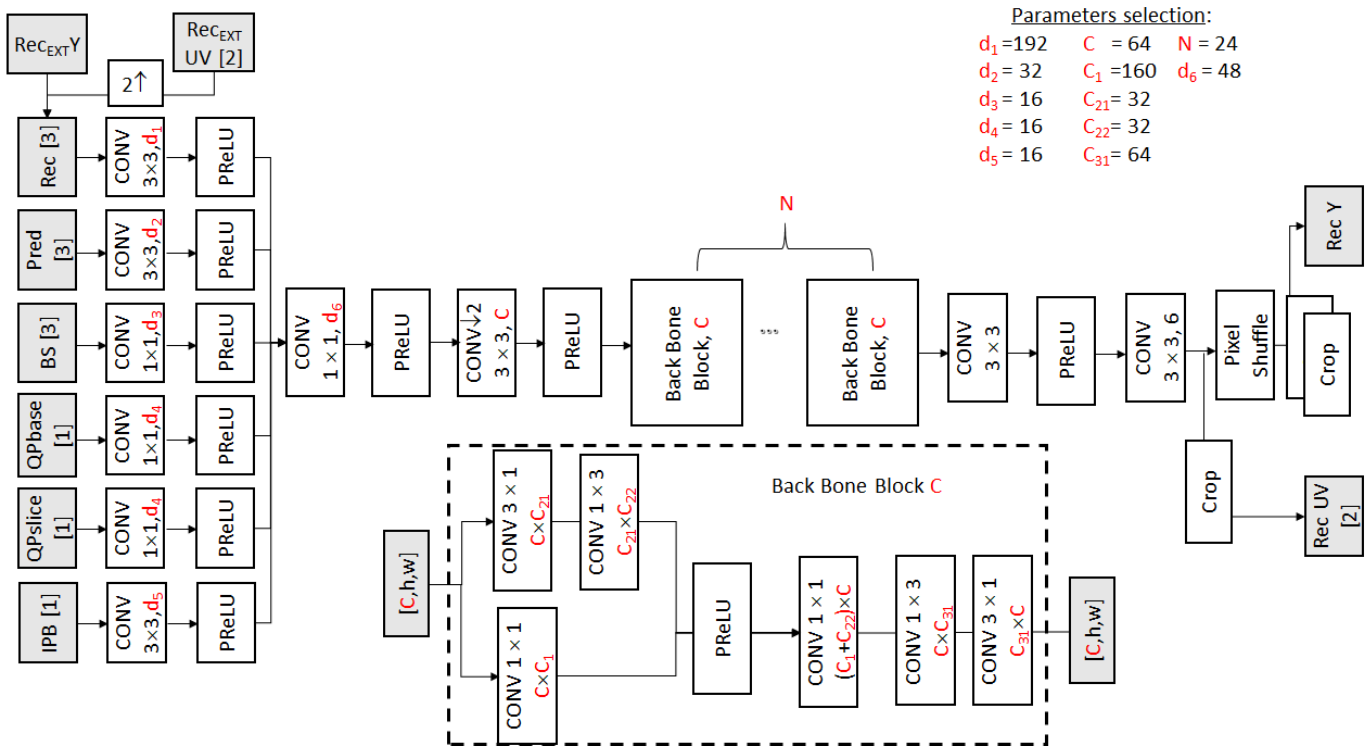
Training conducted: Qualcomm

Cross-check: Inference and training supposed to be cross-checked by Byetedance, cross-check is missed.

This test targets to optimize the complexity-performance trade-off of the convolution decomposition and residual block structures. Reduced rank decomposition and decomposition order alternation of JVET-AE0164 to be tested. The test will include sub-tests with across architecture complexity-performance optimization.

EE1-1.1.3

Input contribution: JVET-AF0103 **EE1-1.1.3 Study on input feature set optimization,** Y. Li, M. Karczewicz (Qualcomm)



Filter architecture: HOP-based

Training strategy: HOP

Training conducted: Qualcomm

Cross-check: Training cross-checked by InterDigital (JVET-AF0258)

This test targets to optimize input feature set extraction, along methods of EE1-1.3.2 (JVET-AD0205). In particular, the headblock design optimization toward improving performance-complexity trade-off and hardware friendliness.

EE1-1.1.4

Input contribution(s):

CITATION

E. Alshina, F. Galpin, Y. Li, D. Rusanovskyy, M. Santamaria, J. Ström, L. Wang, Z. Xie, "EE1: Summary report of exploration experiment on neural network-based video coding", JVET-AF0023, 32nd Meeting: Hannover, DE, October 2023.