|  |
| --- |
| **INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ORGANISATION INTERNATIONALE DE NORMALISATION ISO/IEC JTC 1/SC 29/WG 5 MPEG JOINT VIDEO EXPERTS TEAM WITH ITU-T SG 16** |
| **ISO/IEC JTC 1 / SC 29 / WG 5 N 0291** |
| **Rennes, FR – 17–24 April 2024** |
| |  |  | | --- | --- | | **Title:** | **Exploration experiment on neural network-based video coding (EE1)** | | **Source:** | **Convenor (Jens-Rainer Ohm)** | | **Type:** | **General** | | **Subtype:** | **Other** | | **Status:** | **Approved** | | **Date:** | **2024-05-21** | | **Expected Action:** | **Info** | | **Action due date:** | **N/A** | | **No. of pages** | **12** (without this cover page) | | **Email of convenor:** | **ohm @ ient . rwth-aachen . de** | | **Committee URL:** | **https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-5** | |

|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  34th Meeting, Rennes, FR, 14–24 April 2024 | Document: JVET-AH2023-v4 (v5) |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Exploration Experiments on Neural Network-based Video Coding (EE1)** | | |
| *Status:* | Output document to JVET | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | E. Alshina, R. Chang, F. Galpin, Y.Li, D. Rusanovskyy, M. Santamaria, J. [Ström](mailto:jacob.strom@ericsson.com), and Z. Xie | Email: | [elena.alshina@huawei.com](about:blank)  [renjiechang@tencent.com](mailto:renjiechang@tencent.com)  [franck.galpin@interdigital.com](mailto:franck.galpin@interdigital.com)  [yue.li@bytedance.com](mailto:yue.li@bytedance.com)  dmytror@qti.qualcomm.com  [maria.santamaria\_gomez@nokia.com](mailto:maria.santamaria_gomez@nokia.com)  [jacob.strom@ericsson.com](mailto:jacob.strom@ericsson.com)  [xiezhihuang@oppo.com](mailto:xiezhihuang@oppo.com) |
| *Source:* | EE Coordinators | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document summarizes Exploration Experiment 1 (EE1) tests to be performed between the JVET-AH and JVET-AI meetings to evaluate **Neural Network-based Video Coding (**NNVC) technologies, analyze their performances and complexity aspects.

# Introduction

Code base for the test should be NNVC9.0, anchor is default configuration of NNVC-9.0 (NN-Intra and LOP-3 filter enabled). NNVC common test conditions, results and complexity reporting template must be used.

For proposals **in filter, Intra and super-resolution** categories are encouraged to use **existing AhG11 training set**. Proponents in **NN-Inter category** encouraged to **use extended training set**. In all tests training set must be clearly specified.

For tests competing with technologies in NNVC-9 it is recommended to configure the proposed solution to have a complexity close to related NNVC(LOP/HOP), but not exceeding:

* kMAC/pxl of EE1 test ≤ kMAC/pxl NNVC (*must*),
* Number of Parameters EE1 test ≤ Number of Parameters NNVC (*if possible*).

NN architecture provided in this test description should not be changed, beside minor adjustment for parameters (such as channels number) in order to meet recommendation above. Exact parameters settings can be specified by 2nd AhG11/14 teleconference June 14.

# List of tests

This round of EE1 tests will include:

* EE1-1: LOP in-loop filter
  + EE1-1.1 – LOP with trainable DCT
    - [JVET-AH0207](https://jvet-experts.org/doc_end_user/current_document.php?id=14106) EE1 related: LOP input adjustment with trainable input transform [Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]
  + EE1-1.2 – Partial Convolution and Over-Parameterization
    - [JVET-AH0077](https://jvet-experts.org/doc_end_user/current_document.php?id=13959) AHG11: A Low-Complexity Neural Network Loop Filter based on Partial Convolution and Over-Parameterization [A. Li, C. Zhu, L. Luo (UESTC), Y. Huo, Y. Liu (Transsion)]
  + EE1-1.3 – simplified residual groups with attention
    - [JVET-AH0196](https://jvet-experts.org/doc_end_user/current_document.php?id=14095) [AhG11] On Low Complexity Operational Point for In-Loop Filtering [T. Ryder, S. Eadie, Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]
  + EE1-1.4 – content adaptive VLOP
    - [JVET-AH0096](https://jvet-experts.org/doc_end_user/current_document.php?id=13978) EE1-1.4: Content-adaptive loop-filter, [M.Santamaria, R. Yang, F. Cricri, M. M. Hannuksela, D. Bugdayci Sansli, A. Hallapuro, H. R. Tavakoli, J. Lainema, H. Zhang (Nokia)]
    - [JVET-AH0051](https://jvet-experts.org/doc_end_user/current_document.php?id=13933) EE1-5: Study of the NN architecture at Very Low Operational Point [D. Rusanovskyy, Y. Li, M. Karczewicz (Qualcomm), J. Li, Y. Li, C. Lin, K. Zhang, L. Zhang (Bytedance), Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]
  + EE1-1.5 – Joint VLOP model with transformed inputs
    - [JVET-AH0051](https://jvet-experts.org/doc_end_user/current_document.php?id=13933) EE1-5: Study of the NN architecture at Very Low Operational Point [D. Rusanovskyy, Y. Li, M. Karczewicz (Qualcomm), J. Li, Y. Li, C. Lin, K. Zhang, L. Zhang (Bytedance), Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]
    - JVET-AH0080 EE1-1.2: Joint LOP model with inputs transformed [D. Liu, J. Ström, M. Damghanian, P. Wennersten (Ericsson), D. Rusanovskyy, Y. Li, M. Karczewicz (Qualcomm), T. Shao, P. Yin, S. McCarthy (Dolby), J. N. Shingala, A. Shyam, A. Suneja, S. P. Badya (Ittiam)]
* EE1-2: HOP in-loop filter
  + EE1-2.1 – HOP with transformers and included luma/chroma balance
    - JVET-AH0205, EE1-2.3: Integer implementation of HOP In-loop filter with Transformer blocks and Attention blocks [Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]
    - [JVET-AH0206](https://jvet-experts.org/doc_end_user/current_document.php?id=14105) EE1 related: Additional inference test for EE1-2.3 to adjust luma-chroma balance [Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]
* EE1-3: NN-inter prediction
  + EE1-3.1 – NN-Inter trained with extended training set
    - [JVET-AH0107](https://jvet-experts.org/doc_end_user/current_document.php?id=13989) EE1-3.1: Deep Reference Frame Generation for Inter Prediction Enhancement [W. Bao, N. Fu, X. Chen, J. Jia, Z. Chen (Wuhan Univ.), Z. Liu, X. Xu, S. Liu (Tencent)]
* EE1-4: NN-super-resolution
  + EE1-4.1 – NNSR multiple scaling ratios
    - [JVET-AH0100](https://jvet-experts.org/doc_end_user/current_document.php?id=13982) EE1-related: Multiple scaling ratios coding for NNVC with NNSR [Z. Lv, C. Zhou (vivo)]
  + EE1-4.2 – NNSR with improved training
    - [JVET-AH0312](https://jvet-experts.org/doc_end_user/current_document.php?id=14212) EE1-4 related: On training data of super resolution [C. Lin, Y. Li, J. Li, K. Zhang, L. Zhang (Bytedance)] [late]
* EE1-5. Neural network based Intra coding
  + EE1-5.1 – NN-Intra with ISP
    - [JVET-AH0165](https://jvet-experts.org/doc_end_user/current_document.php?id=14064) AHG11: Combination of the neural network-based intra prediction mode and ISP [T. Dumas, F. Galpin, P. Bordes (InterDigital)]

# Tests description

## EE1-1: LOP in-loop filter

### Test EE1-1.1. [JVET-AH0207](https://jvet-experts.org/doc_end_user/current_document.php?id=14106) EE1 related: LOP input adjustment with trainable input transform [Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

A unified filter architecture for the Low-performance Operation Point (LOP) was defined in JVET-AF0043. Updated training strategy and configuration were presented in JVET-AH0042. Complete inference results for LOP training are reported in JVET-AH0014.

Based on the architecture of the LOP network, input transform with DCT were introduced in JVET-AG0069 and this model has improved the coding gain for RA luma over -0.4 percent compared to LOP with a new two-stage data extraction. This contribution aims at improving the coding gain with a trainable input transform and with the training on the existing LOP2 dataset.

It was commented that this would also be a reasonable extension on top of the new LOP model from EE1-1.2 which uses a DCT.

In this EE test, the DCT component will be replaced by the trainable input transform on top of the LOP3 network, and the network architecture will be adjusted accordingly. The configuration and parameters of the network are subject to further optimization.

Training: Following LOP training strategy, Stage 3 LOP dataset.

Inference: NNVC 9.0 software or latter, NNVC CTC, following LOP anchor settings.

Proponent: Qualcomm

Crosscheck: TBD.

### Test EE1-1.2. [JVET-AH0077](https://jvet-experts.org/doc_end_user/current_document.php?id=13959) AHG11: A Low-Complexity Neural Network Loop Filter based on Partial Convolution and Over-Parameterization [A. Li, C. Zhu, L. Luo (UESTC), Y. Huo, Y. Liu (Transsion)]

This test will include three subtests to assess the effects of partial convolution and over-parameterization:

* EE1-1.2.1: training with architecture changed (only apply over-parameterization).
* EE1-1.2.2: training with architecture changed (only apply partial convolution).
* EE1-1.2.3: training with architecture changed (apply partial convolution and over-parameterization).

The three candidate structures are depicted in Figures EE1-1.2.1/2/3, respectively.

As shown in Figure EE1-1.2.1, we evaluate the effectiveness of the over-parameterization block (OP Block), which expands the model during the training stage while compressing the model during the inference stage.

Training Stage:

As shown in Figure EE1-1.2.1, for a given input feature , where C1 is the number of channels, is the feature height, and is the feature width, the sizes of the two convolution kernels are and , respectively. The convolution process can be expressed as:

,

where ∗ represents the convolution process, and .

Inference Stage:

By the property of convolution, the above process can be re-expressed as:

,

where indicates the fused weight. Through the above transformation, it is possible to use a new convolution kernel to replace the two previous kernels, thereby completing the over-parameterization process.

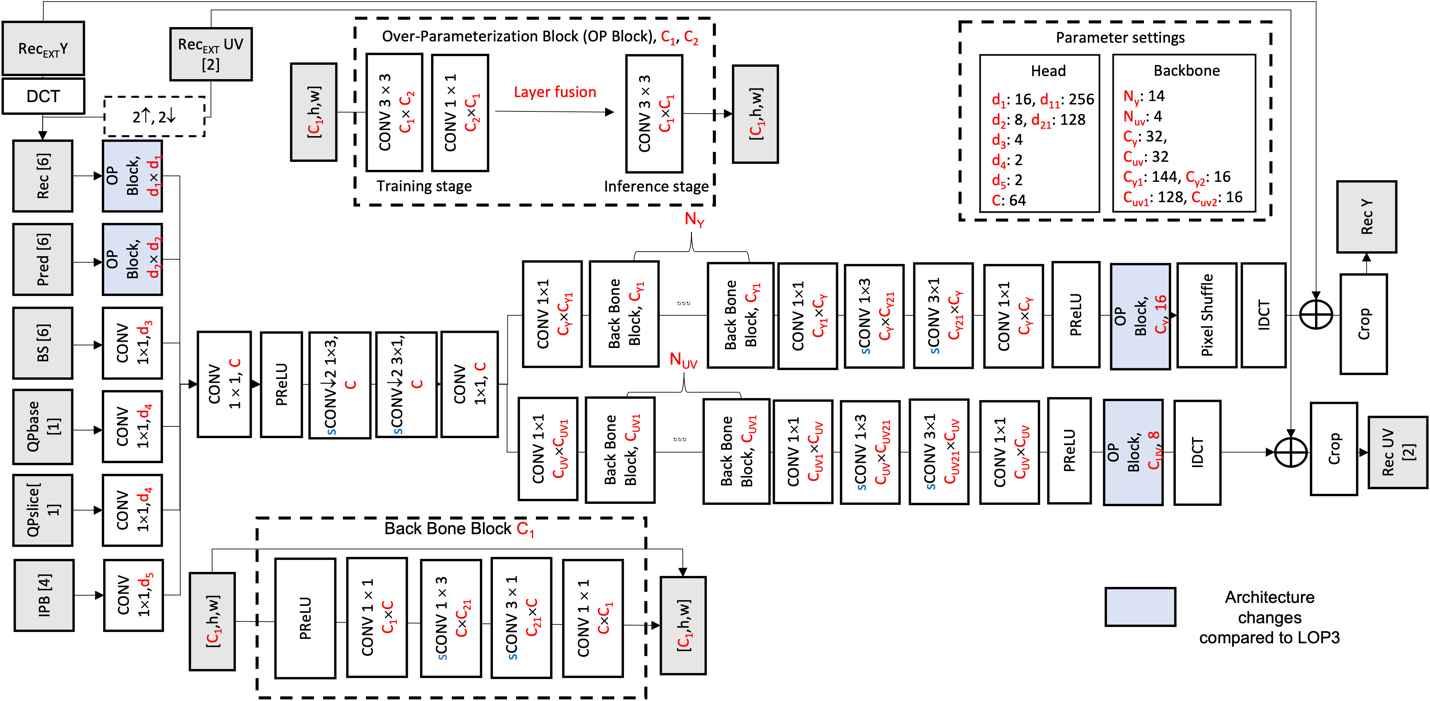


Figure EE1-1.2.1: LOP3 with Over-Parameterization Blocks.

As shown in Figure EE1-1.2.2, we evaluate the effectiveness of the Partial Convolution Block (PC Block), which splits the input features into two branches and performs the convolution on only one of them.

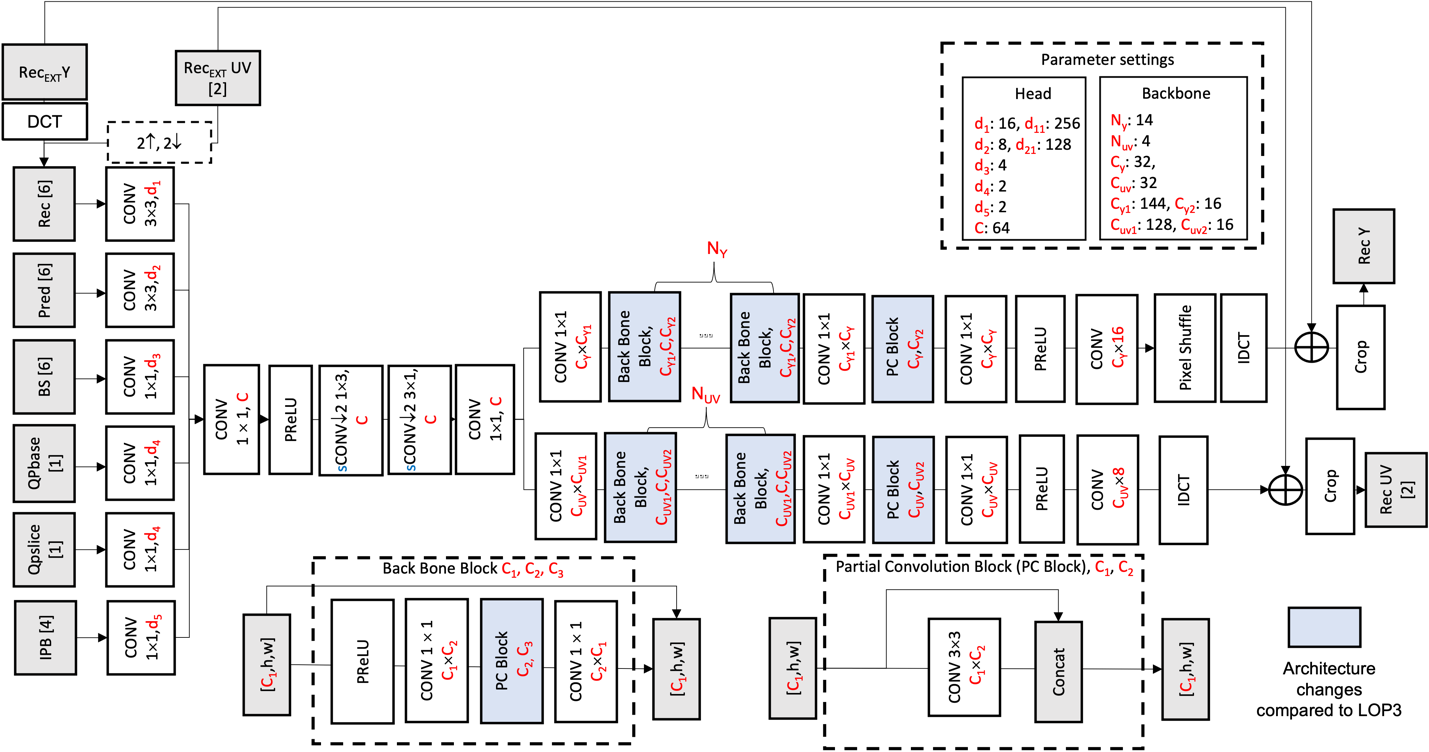


Figure EE1-1.2.2: LOP3 with Partial Convolution Blocks.

As shown in EE1-1.2.3, we combine the Over-Parameterization Block (OP Block) and Partial Convolution Block (PC Block) to evaluate the overall performance."

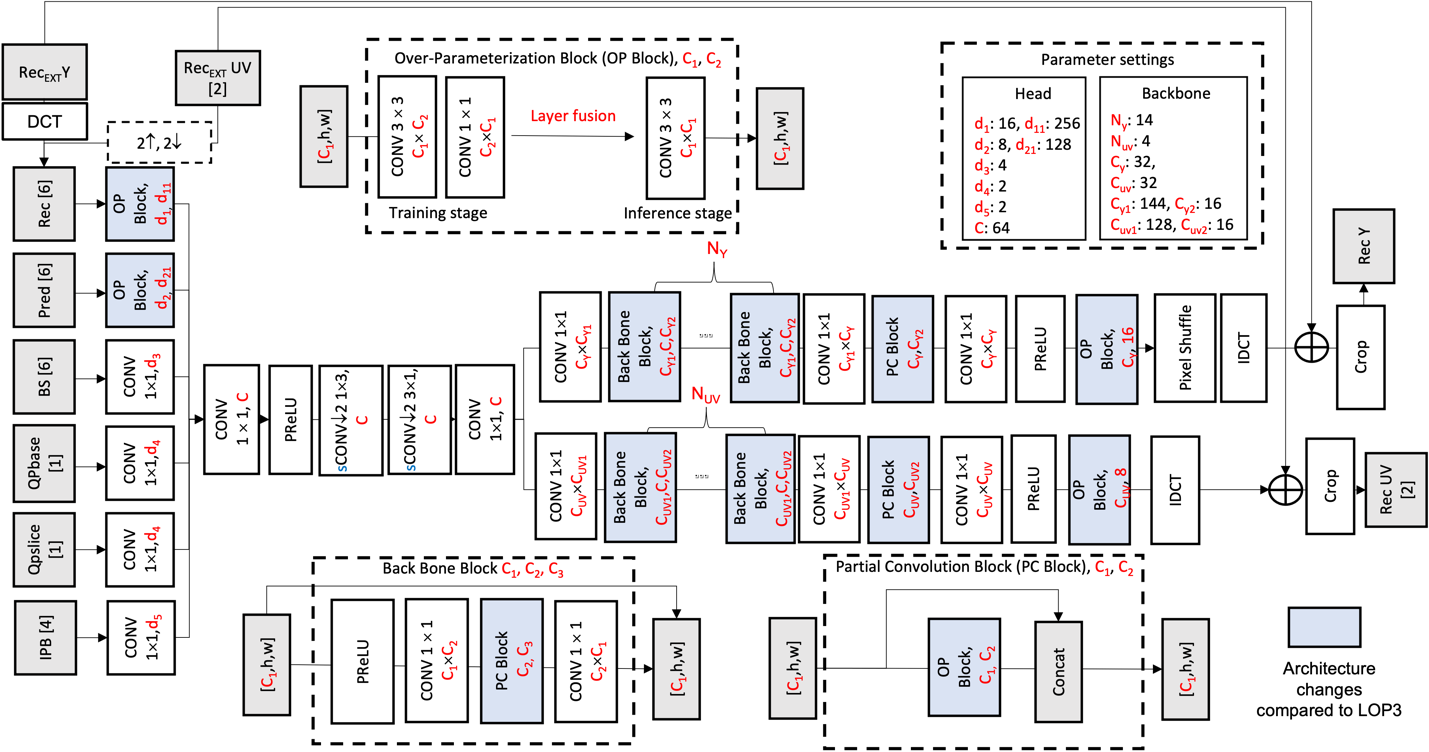


Figure EE1-1.2.3: LOP3 with OP Blocks and PC Blocks.

Training will follow the LOP-3 training strategy, using the Stage 3 LOP dataset. The training process may be modified to reflect changes in the architecture and improve performance. Parameter settings may be adjusted to meet the requirements of the LOP structure.

Training: Use the LOP-3 training strategy and the Stage 3 LOP dataset.

Inference: Employ NNVC 9.0 software or later versions, and NNVC CTC, following the LOP-3 anchor settings.

Proponent: UESTC and Transsion

Crosscheck: TBD.

### Test EE1-1.3. [JVET-AH0196](https://jvet-experts.org/doc_end_user/current_document.php?id=14095) [AhG11] On Low Complexity Operational Point for In-Loop Filtering [T. Ryder, S. Eadie, Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

In this test, a study on NN architecture proposed in JVET-AH0196 for LOP and VLOP is conducted. A candidate architecture for this test is shown in Figure EE1-3.a. It comprises utilization of unified BB across filter architecture and use of Residual Groups with Attention Block, shown in Figure EE1-3.b. The architecture to be modified to achieve target complexity levels for LOP and VLOP and better performance/complexity trade-off. Preliminary parameters of the NN architecture are shown below:

LOP: { d1=16, d2=8, d3=4, d4=d5=2, Cy=28, Cuv=14, Ny=4, Nuv=2, nBBy=4, nBBuv=2}.

VLOP: { d1=4, d2=2, d3=1, d4=d5=1, Cy=16, Cuv=8, Ny=4, Nuv=2, nBBy=4, nBBuv=2}.

A screenshot of a computer

Description automatically generated A black background with white rectangles and red text

Description automatically generated

a. b.

Figure EE1-1.3 Candidate EE1-1.3 architecture (a) and Residual Group (b).

Training to follow LOP training strategy and Stage 3 LOP dataset to be used. Training process may be modified to reflect change in the architecture and improve performance.

Training: Following LOP training strategy, Stage 3 LOP dataset.

Inference: NNVC 9.0 software or latter, NNVC CTC, following LOP and VLOP anchor settings.

Proponent: Qualcomm

Crosscheck: TBD.

### Test EE1-1.4. Content Adaptive Loop-Filter for Very Low Operational Point

The approach of neural network weights adaptation, originally introduced in JVET-AH0096 for LOP architecture, will be applied to Very Low Operational Point filter. A set of multiplier parameters are overfitted to the coded test material. The generated parameter updates are coded with the MPEG-NNR standard and signalled within a NNFU APS.

Training: modified overfitting process adapted for VLOP architecture.

Inference: The NNVC-9.0 (supporting NNFU APS and other changes), NNVC CTC.

Proponent: Joint test of Nokia, Qualcomm

Crosscheck of overfitting and inference is performed by Ericsson.

### Test EE1-1.5. Investigate combination of the modifications from JVET-AH0080 with the VLOP solution of JVET-AH0051

In this test, the VLOP solution from JVET-AH0051 and the modifications of JVET-AH0080 will be studied in combination.

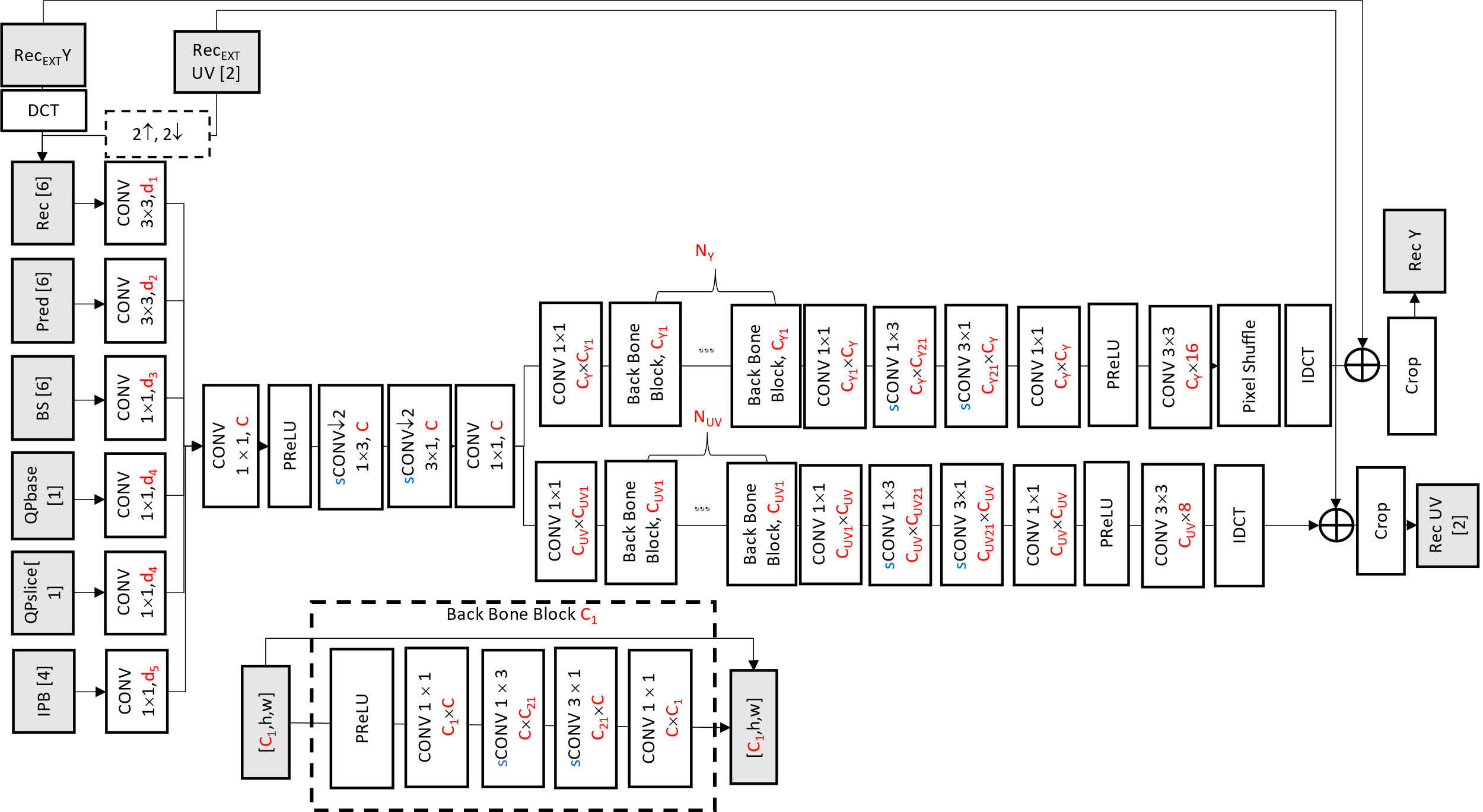


Figure EE1-1.5 Filter architecture of VLOP with transformed input.

Training: Following LOP training strategy, Stage 3 LOP dataset.

Inference: The NNVC-9.0, NNVC CTC.

Proponent: Joint test of Ericsson, Qualcomm

Crosscheck: Ericsson and Qualcomm will independently train/crosscheck each other for the training and inference part.

## EE1-2: HOP in-loop filter

### Test EE1-2.1 [JVET-AH0206](https://jvet-experts.org/doc_end_user/current_document.php?id=14105) EE1 related: Additional inference test for EE1-2.3 to adjust luma-chroma balance [Y. Li, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

A unified filter architecture for the High-performance Operation Point (HOP) was defined in JVET-AD0380. Updated training strategy and configuration were presented in JVET-AG0041, and group convolution, rank reduction and flipping were adopted into HOP in JVET-AG0174. Complete inference results for HOP training are reported in JVET-AH0014.

Based on the architecture of the HOP network, Transformer blocks were introduced in JVET-AF0158, and integer implementation of the Transformer model was described in JVET-AH0205. This model has improved the coding gain significantly compared to HOP. This contribution aims to rebalance the luma-chroma coding gain during inference by adjusting chroma offset.

The EE test consists of two subtests, for the subtest 1, the issues with the complexity assessment for the full transformer model will be further addressed. The configuration and parameters of the network is subject to further optimization.

For the subtest 2, the luma-chroma balance will be adjusted through training and with chroma offset. Other techniques of adjusting the luma-chroma balance may also be introduced in this test.

## EE1-3: NN-inter prediction

### Test EE1-3.1 [JVET-AH0107](https://jvet-experts.org/doc_end_user/current_document.php?id=13989) EE1-3.1: Deep Reference Frame Generation for Inter Prediction Enhancement [W. Bao, N. Fu, X. Chen, J. Jia, Z. Chen (Wuhan Univ.), Z. Liu, X. Xu, S. Liu (Tencent)]

This test is based on JVET-AH0107, aiming at network quantization and training dataset extension. Specifically, this test will implement network quantization, to align the coding performance between float inference and integer inference. Additionally, this test will continue to extend the existing training dataset under NNVC standardization.

The test will be on top of the low operation point deep reference frame generation (LDRF) and high operation point deep reference frame generation (HDRF). The network architectures of both LDRF and HDRF are shown in Figure EE1.3.1

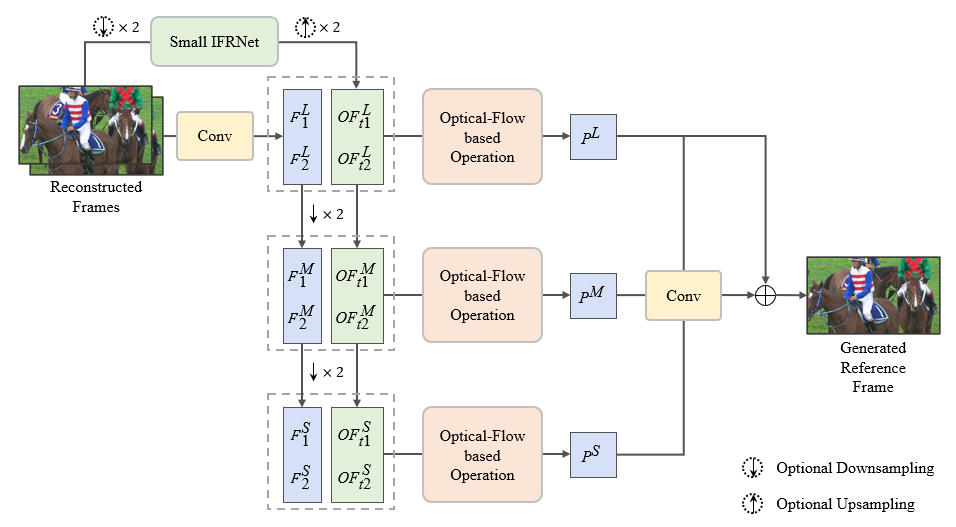


Figure EE1-3.1 Network architecture of EE1-3.1

Training: Training strategy from JVET-AH0107.

Inference: The NNVC 9.0 software is used for inference under NNVC RA configuration.

Proponent: Wuhan University and Tencent.

Cross-check: Inference and training cross-checks are planned.

Cross-checker: OPPO.

## EE1-4: NN-super-resolution

### Test EE1-4.1 [JVET-AH0100](https://jvet-experts.org/doc_end_user/current_document.php?id=13982) EE1-related: Multiple scaling ratios coding for NNVC with NNSR [Z. Lv, C. Zhou (vivo)]

This test aims to investigate a more generic and flexible method for NNVC. Two subtests will be conducted:

* EE1-4.1.1: Adaptive resolution strategy to support multi-scaling ratios (s=2 && s=1.5) with RPR.
* EE1-4.1.2: Unified NNSR for adaptive resolution strategy (EE1-4.1.1) to support multi-scaling ratios (s=2 && s=1.5).

A candidate architecture is shown in Figure EE1-4.1, while the parameters may be changed to reduce complexity. In addition, loss function will be modified by considering subjective quality (introducing SSIM in the loss metric, for example).

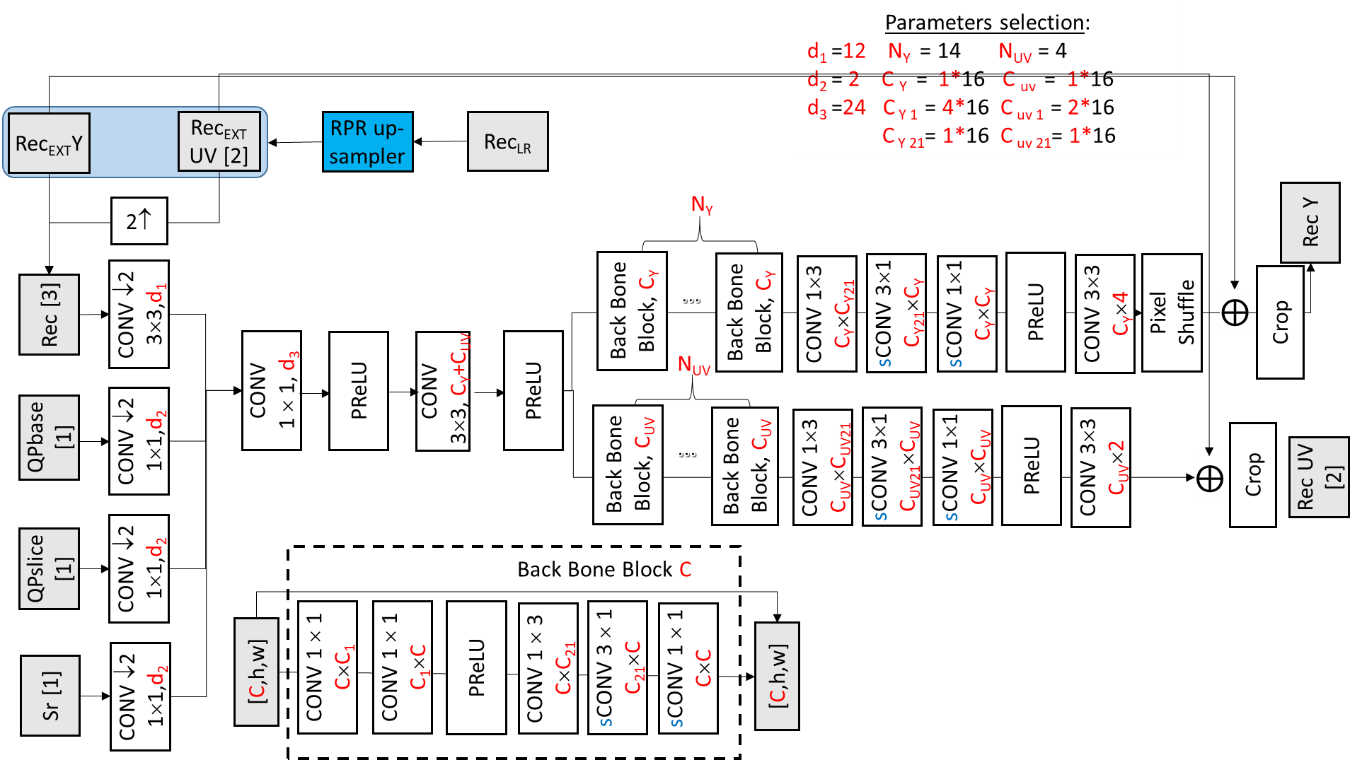


Figure EE1-4.1 Figure x Network structure of EE1-4.1

Inference: The NNVC 9.0 software is used for inference.

Proponent: vivo.

Cross-check: Inference and training cross-checks are planned.

Cross-checker: Bytedance.

### Test EE1-4.2 [JVET-AH0312](https://jvet-experts.org/doc_end_user/current_document.php?id=14212) EE1-4 related: On training data of super resolution [C. Lin, Y. Li, J. Li, K. Zhang, L. Zhang (Bytedance)] [late]

In this test, a study on low complexity super resolution will be conducted. The candidate network architecture and preliminary parameters for this test are shown in Figure EE1-4.2. The network architecture and parameters may be modified to achieve better performance/complexity trade-off. Additionally, training parameters and loss function may be modified to improve coding performance and subjective quality.



Pixel Shuffle

Rec Y

Rec UV [2]

CONV 1 × 1 C×C

Figure EE1-4.2 Network structure of EE1-4.2

Inference: The NNVC 9.0 software is used for inference.

Proponent: Bytedance.

Cross-check: Inference and training cross-checks are planned.

Cross-checker: vivo.

## EE1-5 Neural network based Intra coding

### Test EE1-5.1 [JVET-AH0165](https://jvet-experts.org/doc_end_user/current_document.php?id=14064) AHG11: Combination of the neural network-based intra prediction mode and ISP [T. Dumas, F. Galpin, P. Bordes (InterDigital)]

This contribution proposes a combination of the Neural Network (NN)-based intra prediction mode and ISP inside NNVC. In addition, the objective function for training a NN for intra prediction now involves the SATD and the SAD between a given original block and its neural network prediction as this objective function has yielded mean BD-rate gains when combining the NN-based intra prediction mode and ISP.

NNVC-8.0 involving LOP-2, the low-complexity NN-based intra prediction mode, and the proposed combination of the NN-based intra prediction mode and ISP with respect to NNVC-8.0 involving LOP-2 and the default NN-based intra prediction mode (NNVC-8.0 CTC) gives:

* RA: -0.13%, -0.78% and -0.76% BD-rate PSNR Y, U and V gains, respectively for a complexity of 112% and 113% at encoder and decoder respectively
* AI: -0.17%, -0.63% and -0.56% BD-rate PSNR Y, U and V gains, respectively, for a complexity of 155% and 115% at encoder and decoder respectively.

NNVC-8.0 involving LOP-2, the low-complexity NN-based intra prediction mode, and the proposed combination of the NN-based intra prediction mode and ISP with respect to NNVC-8.0 involving LOP-2 and the low-complexity NN-based intra prediction mode provides:

* RA: -0.25%, -0.85% and -0.80% BD-rate PSNR Y, U and V gains, respectively for a complexity of 107% and 103% at encoder and decoder respectively
* AI: -0.54%, -0.93% and -0.93% BD-rate PSNR Y, U and V gains, respectively, for a complexity of 151% and 109% at encoder and decoder respectively.

Number of operations is decreased relative to current NN intra pred., but decoder runtime is increased due to more frequent usage.

Interesting gain and reduction in computation, but encoder runtime increased significantly, due to necessity of more checks.

In this EE test the ocus should be reduction of encoding time.

Diagram

Description automatically generated

Figure EE1-5.1.1 prediction of the current block from its context of reference samples via the neural network , parametrized by , belonging to the low-complexity NN-based intra prediction mode. Here, and The definitions of , , , “preprocessing”, and “postprocessing” can be found in JVET-AG2019.

Table 1: architecture of each neural network belonging to the low-complexity NN-based intra prediction mode.

|  |  |  |
| --- | --- | --- |
| **architecture** | | **description** |
| Number of hidden layers | | For 16x16, 3.  For the other six NNs, 2. |
| Number of neurons per hidden layer | | For 4x4, 8x4, and 16x16, 1216 for the last hidden layer, and 640 for the other hidden layers.  For the other four NNs, 1216. |
| Percentage of non-zero weights | hidden layers | 5% |
| layer returning | For 4x4, 10%. For the other six NNs, 50% |
| layer returning the predicted block | 50% |
| layer returning and | 100% |

Worst-case complexity of the low-complexity NN-based intra prediction mode: 4.8 kMAC/pixel.

Number of parameters of this mode (7 NNs together): 1333799.

# Timeline

**T1 - 2 weeks after JVET-AH meeting (10-May-2024):** EE description (JVET-AH2023) finalized and uploaded.NNVC-8.0 software is available, including anchor performance.

**T2 – 3 weeks after JVET-AH meeting (17-May-2024) 1st teleconference.** HOP training is over. Results discussed during teleconference.

**T2 – 7 weeks after JVET-AH meeting (14-June-2024) 2nd teleconference.** Final setting for parameters to be announced, partial results discussed, combinational tests (if any) decided.

**T3 – 2 weeks before T5 (27-June-2024)**: EE1 software is frozen (write access closed on 29-Jun-2024).

**T4 – 3 days before T5 (10-July-2024):** Cross-checkers report status to EE1 coordinators (sending e-mail).

**T5 –10-July-2024:** EE1 summary is uploaded as input contribution.

# SW location

https://vcgit.hhi.fraunhofer.de/jvet-ah-ee1/VVCSoftware\_VTM/-/tree/EE1-X.X

# References

[1] E. Alshina, R.-L. Liao, S. Liu, A. Segall Common test conditions and evaluation procedures for neural network-based video coding technology, [JVET-AG2016](https://jvet-experts.org/doc_end_user/current_document.php?id=13272).

[2] E. Alshina, F. Galpin BoG on NNVC training material [JVET-AH0354](https://jvet-experts.org/doc_end_user/current_document.php?id=14254)