|  |
| --- |
| **INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ORGANISATION INTERNATIONALE DE NORMALISATION ISO/IEC JTC 1/SC 29/WG 5 MPEG JOINT VIDEO CODING TEAM WITH ITU-T SG 16** |
| **ISO/IEC JTC 1 / SC 29 / WG 5 N 113** |
| **Online, 12–21 January 2022** |
| |  |  | | --- | --- | | **Title:** | **Exploration experiment on neural network-based video coding (EE1)** | | **Source:** | **Convenor (Jens-Rainer Ohm)** | | **Type:** | **General** | | **Subtype:** | **N/A** | | **Status:** | **Approved** | | **Date:** | **2022-02-04** | | **Expected Action:** | **Info** | | **Action due date:** | **N/A** | | **No. of pages** | **8** (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**  25th Meeting, by teleconference, 12–21 January 2022 | Document: JVET-Y2023\_r1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Exploration Experiments on Neural Network-based Video Coding (EE1)** | | |
| *Status:* | Output Document to JVET | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | E. Alshina, W. Chen, F. Galpin, Y. Li, Z. Ma, H. Wang, and L.Wang | Tel: Email: | [elena.alshina@huawei.com](about:blank),  [chenwei06@kwai.com](about:blank),  Franck.Galpin@InterDigital.com,  yue.li@bytedance.com, [mazhan@nju.edu.cn](about:blank),  [hongtaow@qti.qualcomm.com](about:blank)  liqiangwang@tencent.com |
| *Source:* | EE coordinators | | |

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

# Abstract

This document describes Exploration Experiments (EEs) planned to be performed between the JVET-Y and JVET-Z meetings to evaluate **Neural Network-based Video Coding (**NNVC) technologies, analyze their performance, and analyze their complexity aspects.

# Introduction

Group continues evaluation of new promising NN-based video coding technologies, answering questions and addressing suggestions from JVET members made during presentation NN-based technologies at JVET-Y meeting. Additionally in this round of EE group will focus on investigation platform-independent reproducibility (which is desirable feature for video codec [1]), drift-free loop operation by integerization of NNs, and usage of a software package which supports that.

The ultimate goal of the NNVC Exploration Experiments would be the creation of AhG11 anchor based SW with examples of NN-tools implementation for further collaborative development of NN-based video coding tools.

Tests will be conducted in two categories: enhancement filters and super-resolution methods. NN-based intra prediction was studied in several rounds of EE1; tool is stable, successfully crosschecked, performance verified on top of VTM and ECM. No study for NN-based Intra is planned at this point.

All proponents must use AhG11 anchor [2] (VTM-11.0 + “newMCTF” patch [3], QP=22, 27, 32, 37, 42) and the reported template recommended by AhG11.

Proponents are encouraged to report both CPU and GPU decoding run time.

Test results and complexity analysis reporting template [2] are expected to be uploaded together with final software by the T4 deadline specified in section “Timeline”.

Discussions with regards to this EE are expected to be conducted in JVET reflector.

# Exploration experiments on Enhancement filters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Proposal | Proponent | Contact | Cross-checker |
| 1.1 | [JVET-Y0078](https://jvet-experts.org/doc_end_user/current_document.php?id=11272) | Tencent | liqiangwang@tencent.com | [xiezhihuang@oppo.com](mailto:xiezhihuang@oppo.com) |
| 1.2 | [JVET-Y0080](https://jvet-experts.org/doc_end_user/current_document.php?id=11274) | Tencent | liqiangwang@tencent.com | [xiezhihuang@oppo.com](mailto:xiezhihuang@oppo.com) |
| 1.3 | [JVET-Y0098](https://jvet-experts.org/doc_end_user/current_document.php?id=11292) | Ericsson | [kenneth.r.andersson@ericsson.com](mailto:kenneth.r.andersson@ericsson.com) |  |
| 1.4 | [JVET-Y0046](https://jvet-experts.org/doc_end_user/current_document.php?id=11238) | Xidian University & ZTE | [yi.zhou@stu.xidian.edu.cn](mailto:yi.zhou@stu.xidian.edu.cn) | [linchaoyi.cy@bytedance.com](mailto:linchaoyi.cy@bytedance.com) |
| 1.5 | [JVET-Y0052](https://jvet-experts.org/doc_end_user/current_document.php?id=11246) | Xidian University & OPPO | [13227706628@163.com](mailto:13227706628@163.com) |  |
| 1.6. | [JVET-Y0143](https://jvet-experts.org/doc_end_user/current_document.php?id=11337)  [JVET-Y0110](https://jvet-experts.org/doc_end_user/current_document.php?id=11304) | Bytedance,  Qualcomm, Interdigital | [yue.li@bytedance.com](mailto:yue.li@bytedance.com)  [franck.galpin@interdigital.com](mailto:franck.galpin@interdigital.com) |  |
| 1.7 | [JVET-Y0143](https://jvet-experts.org/doc_end_user/current_document.php?id=11337)  [JVET-Y0098](https://jvet-experts.org/doc_end_user/current_document.php?id=11292)  [JVET-Y0110](https://jvet-experts.org/doc_end_user/current_document.php?id=11304) | Bytedance,  Qualcomm, Ericsson, Interdigital | yue.li@bytedance.com |  |

1.1 [JVET-Y0078](https://jvet-experts.org/doc_end_user/current_document.php?id=11272) EE1-1.1: neural network based in-loop filter with constrained storage and low complexity

A neural network based in-loop filter with constrained parameter memory and low complexity is proposed in this contribution. For saving the parameter memory, there are **only two models** used in the proposed filter design, for which each model is designed with relatively low complexity. For these two models, one is designed for I slices, and the other one is for B slices. The final filtered result is decided between two results from the NN based filter and the traditional filter. Besides, the NN based filter can be turned on/off at both the CTU level and the slice level.

The specific network framework is provided as the following figure. In addition to the reconstruction image, other side information is fed into the network, such as the prediction image, partition image, slice QP and based QP. As for the design spirit of the ResBlock, the number of channels firstly goes up before the activation layer, and then goes down after the activation layer.

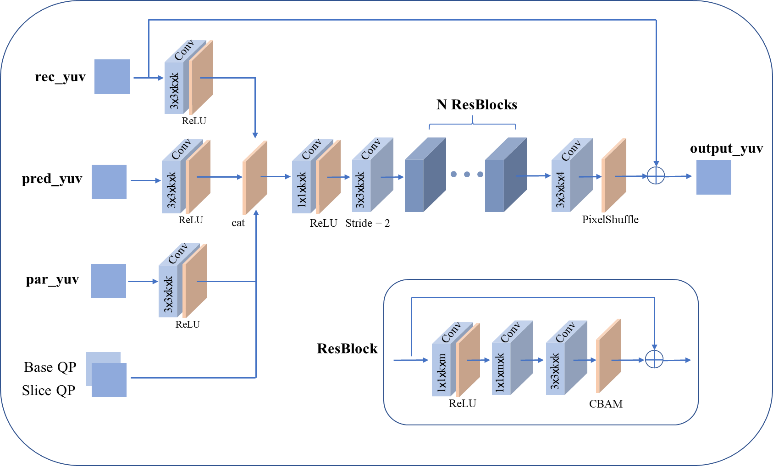


Fig. 1. NN structure of JVET-Y0078.

Test 1.1: Optimize the filter in both the network and the implementation based on JVET-Y0078.

Test 1.1.2: Float SADL implementation of Test 1.1.1.

Test 1.1.3: Integer SADL implementation of Test 1.1.1.

1.2 [JVET-Y0080](https://jvet-experts.org/doc_end_user/current_document.php?id=11274) EE1-1.1-related: alternative filter designs [L. Wang, X. Xu, S. Liu (Tencent)]

This test evaluates the filter 2 in JVET-Y0080, which is a further development for NN-based filter studied in EE before. Compared with JVET-Y0078, the number of models is decreased into only one for further saving the parameter memory. For different base QPs and different color components, there is **only one model** used in the filtering process.

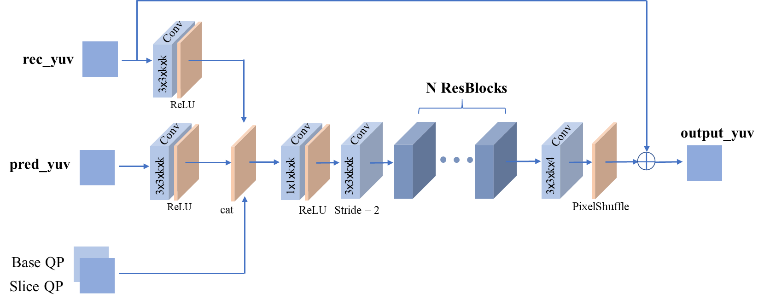


Fig. 2. NN structure of JVET-Y0080.

Test 1.2: Optimize the filter in both the network and the implementation based on JVET-Y0080.

Test 1.2.2: Float SADL implementation of Test 1.2.1.

Test 1.2.3: Integer SADL implementation of Test 1.2.1.

1.3 [JVET-Y0098](https://jvet-experts.org/doc_end_user/current_document.php?id=11292) EE1-related: Combination of VVC deblocking and NN loop filtering [K. Andersson, J. Ström, D. Liu, R. Sjöberg (Ericsson)]

Improvement on top of NN loop filtering filter studied in EE before, solving problematic situation in which no de-blocking operations performed. .

Other differences are as follows. Instead of signalling which QP to use for the model from a fixed set of QPs, only one QP is used. Also, the maximum block size (excluding border extension for input) for luma is restricted to 128x128 for inter pictures for resolutions below 4K. The output is a convex combination of the VVC deblocked samples and the NN-processed samples. Since both VVC deblocking and NN loop filtering provide a deblocking effect, the output samples will always have seen some form of deblocking. The VVC deblocking is performed in RDO and a deblocking beta offset of -2 is also used.

Operation in parallel, and then a weighted combination of NN and deblocking output is used. Weight is optimized by MSE minimization, and signalled at slice level.

Test 1.3: NN-based filter as proposed in JVET-Y0098

1.4 [JVET-Y0046](https://jvet-experts.org/doc_end_user/current_document.php?id=11238) AHG11: ALF improvement for NNVC [W. Zou, Y. Zhou (Xidian Univ.), C. Huang, Y. X. Bai (ZTE)]

This contribution presents an ALF modification in NNVC. When a neural network-based filter is used as in-loop filter, two ALF flags, which indicate ALF enabled/disable, are introduced for luma and chroma components, respectively.

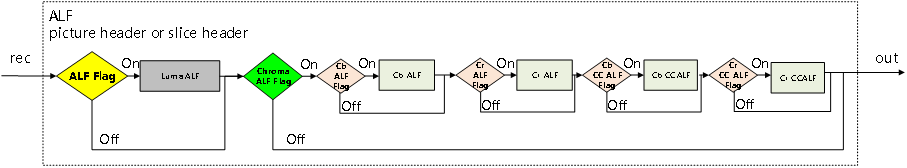


Fig. 3. *Modified Syntax structure*.

Goal of this experiment is to explore ALF\_SPLIT performance under different configurations.

Test 1.4.1: JVET-X0065 anchor with the ALF-SPLIT tool presented in JVET-Y0046.

Test 1.4.2: Test 1.4.1 with larger chroma QP offset.

1.5 [JVET-Y0052](https://jvet-experts.org/doc_end_user/current_document.php?id=11246) AHG11: CNN post-processing filter based on depthwise separable convolution and attention mechanism [H. Zhang, C. Jung (Xidian University), D. Zou, M. Li (OPPO)]

This contribution proposes weakly connected dense attention neural network for compression artifact removal, called WCDANN. WCDANN is a post-processing filter based on convolutional neural networks (CNNs) that does not need any change of codecs. WCDANN consists of several weakly connected dense attention blocks (WCDABs), which is based on residual learning by taking the compressed video after codecs as the input. Depthwise separable convolution are used; as the basic convolution unit for WCDANN to generate a lightweight model. Moreover, we introduce attention mechanism into the proposed filter to emphasize important features.

Two models for lower/higher QP ranges

Training uses L1 loss first, then switches to L2

Only applied on Y.

Test 1.5: NN-based filter as proposed in JVET-Y0053

1.6 [JVET-Y0143](https://jvet-experts.org/doc_end_user/current_document.php?id=11337) EE1-1.2: Test on Deep In-Loop Filter with Adaptive Parameter Selection and Residual Scaling [Y. Li, K. Zhang, L. Zhang (Bytedance), H. Wang, K. Reuzé, A.M. Kotra, M. Karczewicz (Qualcomm)]

[JVET-Y0110](https://jvet-experts.org/doc_end_user/current_document.php?id=11304) AHG11: Small Ad-hoc Deep-Learning Library (SADL) update [F. Galpin, F. Mom, T. Dumas, P. Bordes, P. Nikitin, E. François (InterDigital)]

This test evaluates the effectiveness of the deep in-loop filter presented in JVET-Y0143 re-implanted using SADL library (JVET-Y0110).

Filter design is as follows. First, the proposed filter takes auxiliary information as input. Second, external attention mechanism is introduced. Third, a slice or block could determine whether to apply the CNN-based filter or not. When the CNN-based filter is determined to be applied to a slice/block, the filtering parameter selected from multiple candidates is further decided. At last, when a NN filter is being applied to reconstructed pictures, the difference between the input samples and the NN filtered samples (residues) are scaled by the scaling factors before being added to input samples.



Fig. 4. Neural network structure of JVET-X0143.

Currently filter is implemented using libTorch and float-point arithmetic. Goal of the test to evaluate possibility and associated performance (speed) change due to re-implementation with SADL. Due to the float point arithmetic performance of this implementation still not have bit-exact behaviour. Performance difference between Linux/Windows and drift due to error accumulation will be studied.

If time allows then next step of this study (NN quantization and testing fixed point implementation).

Test 1.6.1: VTM anchor, SADL implementation of JVET-X0143 (float point arithmetic).

Test 1.6.2: VTM anchor, SADL implementation of JVET-X0143 (quantized NN parameters, integer implementation).

1.7 Joint test of EE1-1.6 (JVET-Y0143, JVET-Y0110) and EE1-1.3 (JVET-Y0098)

This test evaluates the combination of test EE1-1.6 and test EE1-1.3.

# Exploration experiments on NN-based super resolution

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Proposal | Proponent | Contact | Cross-checker |
| 2.1 | [JVET-Y0068](https://jvet-experts.org/doc_end_user/current_document.php?id=11262) | LGE | [junghak.nam@lge.com](mailto:junghak.nam@lge.com) | [kenneth.r.andersson@ericsson.com](mailto:kenneth.r.andersson@ericsson.com) |
| 2.2 | [JVET-Y0069](https://jvet-experts.org/doc_end_user/current_document.php?id=11263) | Bytedance | [linchaoyi.cy@bytedance.com](mailto:linchaoyi.cy@bytedance.com) |  |
| 2.3 | [JVET-Y0070](https://jvet-experts.org/doc_end_user/current_document.php?id=11264) | Bytedance | [linchaoyi.cy@bytedance.com](mailto:linchaoyi.cy@bytedance.com) |  |
| 2.4 | [JVET-Y0068](https://jvet-experts.org/doc_end_user/current_document.php?id=11262)  [JVET-Y0070](https://jvet-experts.org/doc_end_user/current_document.php?id=11264) | LGE  Bytedance | [junghak.nam@lge.com](mailto:junghak.nam@lge.com)  [linchaoyi.cy@bytedance.com](mailto:linchaoyi.cy@bytedance.com) |  |

2.1 [JVET-Y0068](https://jvet-experts.org/doc_end_user/current_document.php?id=11262) EE1-2.1-related: RPR encoder with multiple scale factors [J. Nam, S. Yoo, J. Lim, S. Kim (LGE)]

This contribution proposes encoder control for choosing using picture size for encoding from multiple scale factors, such as x2.0 (half size) and x1.5 (2/3 size). At each GOP, the encoder can select a scale factor based on initial QP and a PSNR value which is calculated between the source and down-up sampling pictures for the first picture of each GOP.

Additionally to proposed technology it was suggested by JVET to study adjustment CbQpOffset and CrQpOffset for reducing Chroma loss. To match bitrates, it also study luma QP adjustment along with chroma.

Test 2.1: Adaptive resolution selection w/o multi-pass as proposed in JVET-Y0068.

Test 2.1.2: Test 2.1.1 with luma and chroma QP adjustments.

2.2 [JVET-Y0069](https://jvet-experts.org/doc_end_user/current_document.php?id=11263) EE1-2.3: CNN-based Super Resolution for Video Coding Using Decoded Information [C. Lin, Y. Li, K. Zhang, L. Zhang (Bytedance)]

This test evaluates the effectiveness of the test 2.3.2 proposed in JVET-Y0069, which is a simplified CNN-based super resolution. The decoded information is fed into the up-sampling networks designed for luma and chroma components, respectively.

The up-sampling model for the luma component is fed with: (i) QP, (ii) luma samples of reconstruction, (iii) luma samples of prediction.

The up-sampling model for the chroma component is fed with: (i) luma samples of reconstruction, (ii) QP, (iii) chroma samples of reconstruction，(iv) chroma samples of prediction.



Figure 5. Luma neural network structure in Test 2.2



Figure 6. Chroma neural network structure in Test 2.2

Test 2.2 VTM-11-NNVC anchor with JVET-Y0069

2.3 [JVET-Y0070](https://jvet-experts.org/doc_end_user/current_document.php?id=11264) EE1-2.4: CNN-based Super Resolution for Video Coding Using Separate Networks for Chroma Components [C. Lin, Y. Li, K. Zhang, L. Zhang (Bytedance)]

This test evaluates the effectiveness of the CNN-based super resolution proposed in JVET-Y0070, which uses separate models for up-sampling the Y, U, and V components, respectively. The decoded information is fed into the up-sampling networks designed for luma and chroma components, respectively.

The up-sampling model for the luma component is fed with: (i) QP, (ii) luma samples of reconstruction, (iii) luma samples of prediction.

The up-sampling model for the chroma component is fed with: (i) luma samples of reconstruction, (ii) QP, (iii) chroma samples of reconstruction，(iv) chroma samples of prediction.



Figure 7. Luma neural network structure in Test 2.3



1. Network structure for up-sampling U components.



1. Network structure for up-sampling V components.

Figure 8. Chroma neural network structure in Test 2.3

Test 2.3: VTM-11-NNVC anchor with JVET-Y0070

2.4 Joint test of EE1-2.1.2 ([JVET-Y0068](https://jvet-experts.org/doc_end_user/current_document.php?id=11262)) and EE1-2.3 ( [JVET-Y0070](https://jvet-experts.org/doc_end_user/current_document.php?id=11264))

Best variant of adaptive resolution found in test 2.1.2 (includes luma / chroma QP adjustment) with be tested with NN-based up-sampling filter from test 2.3.

Test 2.4: replace RPR up-sampling filter in Test 2.1.2 by NN-based up-sampling from Test 2.3.

# Visual test

EE1 tests which are recommended for viewing, test sequences and QPs for anchor and adaptive resolution (Test 2.1) coding will be selected at teleconference (T3 deadline specified in section “Timeline”).

During telco proponents can provide suggestions which video sequences to be included to the viewing. Candidates recommended for the viewing requested to prepare and upload mp4, using naming convention and instructions below:

1. Use the latest version of ffmpeg.
2. Select QP for coding in order to ensure **no larger than 10% difference** between bit stream sizes of AhG11/EE1 anchor and test.
3. Use following sequence of commands for conversion decoded yuv (10 bits 4:2:0) to mp4:

ffmpeg -s:v [W]x[H] -c:v rawvideo -pix\_fmt yuv420p10le -i [SEQUENCE\_NAME]\_QP[QP\_VALUE].yuv -filter:v select="between(n\, [START]\, [END])" temp.yuv

ffmpeg -s:v [W]x[H] -c:v rawvideo -r [FRAME\_RATE] -pix\_fmt yuv420p10le -i temp.yuv -c:v libx265 -crf 15 -tag:v hvc1 -pix\_fmt yuv420p10le [SEQUENCE\_NAME]\_QP[QP\_VALUE].mp4

Parameters [SEQUENCE\_NAME], [W],[H], [QP\_VALUE], [START], [END] will be decided during Ahg11/EE1 telco (T3 deadline specified in section “Timeline”) and send to the proponents.

In file name [QP\_VALUE] is QP which was used for coding (might be different from anchor); [SEQUENCE\_NAME] should match sequence name in \cfg\per-sequence. Parameters [START], [END] specify start and end frames of video sequences (only 5 sec segments will be used during viewing).

1. Upload prepared mp4 to the folder wit experiment name

[NAME\_OF\_EXPERIMENT]/[SEQUENCE\_NAME]\_QP[QP\_VALUE].mp4

For example, results of coding DaylightRoad2\_3840x2160\_60fps\_10bit\_420.yuv sequence during test EE1-2.4 to be compared with AhG1 anchor coded at QP=42 should be EE1-2.4/DaylightRoad2\_QP42.mp4; results of coding Campfire\_3840x2160\_30fps\_10bit\_420\_bt709\_videoRange.yuv sequence during test EE1-1.1.3 to be compared with AhG1 anchor coded at QP=37 should be EE1-1.1.3/ Campfire\_QP37.mp4.

Remote viewing will be conducted in coordination with AG5 at (T4 deadline specified in section “Timeline”). List of sequences, QPs and updated (if needed) instruction with naming convention and examples of command lines for viewing preparation will be released shortly after T3.

Viewing sessions will be announced via JVET reflector and added to JVET calendar.

# Timeline

**T1 - 2 week after JVET-Y meeting (04-Feb-2022):** To revise EE description. Changes should be discussed and agreed on JVET reflector. Anchor is available.

**T2 – 3 weeks after JVET-Y meeting (11-Feb -2022)**: Initial software release (which includes training scripts for proposals supposed to have training verified) that matches what was proposed to the meeting.

**T3 – 4 weeks after JVET-Y meeting (18-Feb -2022)**: Teleconference for viewing preparation. AhG11 suggests sequences and QP points for anchor and RPR.

**T4 - ~4 weeks before T6 (21-March-2022):** Software is frozen (and may include improvements), technology description is ready, and cross-check starts. Remote viewing session conducted.

**T5 – 3 days before T6 (18-April-2022):** Cross-checkers report status to EE coordinators.

**T6 - 21-April-2022:** EE summary is uploaded as input contribution.

# . References

[1] BoG Report: Neural Networks Video Coding Analysis and Planning, JVET-W0182.

[2] Common Test Conditions and evaluation procedures for neural network-based video coding technology, JVET-W2016.

[3] <https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc/-/blob/master/Software%20Patches/JVET-V0056_VTM11.0_backport.patch>