 ISO/IEC JTC 1/SC 29/WG 5 N0057

**ISO/IEC JTC 1/SC 29/WG 5**

**MPEG Joint Video Coding Team(s) with ITU-T SG 16**

**Convenorship: DE**

**Document type:** General

**Title:** Exploration experiment on neural network-based video coding (EE1)

**Status:** Approved

**Date of document:** 2021-04-28

**Source:** ISO/IEC JTC 1/SC 29/WG 5

**Expected action:** Info

**Action due date:** None

**No. of pages:** 6 (without this cover page)

**Email of Convenor:** ohm @ ient . rwth-aachen . de

**Committee URL:** https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg5

|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  22nd Meeting, by teleconference, 20–28 Apr. 2021 | Document: JVET-V2023\_r1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Exploration Experiments on Neural Network-based Video Coding** | | |
| *Status:* | Output document of JVET | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | E. Alshina, S. Liu, W. Chen, F. Galpin, Y. Li, Z. Ma and H. Wang | Email: | [elena.alshina@huawei.com](about:blank), [shanl@tencent.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) |
| *Source:* | EE coordinators | | |

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

This document describes Exploration Experiments (EEs) planned to be performed between JVET-V and JVET-W meetings in order to evaluate NNVC technologies, analyze their performance and complexity aspects.

# Introduction

In the 22nd JVET meeting it was decided to continue an exploration experiment on NN-based video coding and include technologies listed in section 2. The goal of EE at this stage includes:

* More comprehensive visual quality analysis
* Cross-check of available data (including complexity)

Discussion of EE description during editing period supposed to happen in main JVET reflector. SW for each test will be made available for all JVET members with proper announcement in main JVET reflector. EE participants agreed that proponent will provide instruction for SW installation and usage and assist JVET members in case they meet trouble with running the code.

Proponents must follow NN-CTC [1], use results reported template and fill out complexity assessment page (at least mandatory elements as requested in [1]). Namely VTM11.0 with new MCTF ([JVET-V0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10704)) [2] should be used as anchor, GOPsize in RA is 32, LDP and LDB configuration for UHD sequences is optional (sequence length is 3 sec for UHD content in LDB and LDP).

Proponents are mandated to report MAC/pxl computation using methodology described in [1]. Example of MAX/pxl computation will be available in [4]. Cross-checkers should verify not only compression performance, but also reported complexity to be aligned with NN structure described in proposal.

Results Reporting Template [1] must be used by all EE proponents.

Proponents and cross-checkers are encouraged to test technology using materials out of NN-CTC [1] and report their observation to the group.

Detailed information about training beyond mandatory in [1] (such as training sub-set selection logic, learning curve) is encouraged to be provided by proponent.

Proponents are highly encouraged to freeze the models by the deadline in order to make the crosscheck and viewing preparation by AhG11 possible .

There are two categories of tests: NN based in-loop filters and super resolution technologies.

# List of experiments

## NN-based filtering

The majority of proponents in this category decided not to change filter design compare to JVET-V-EE1. Only code modification related to new MCTF ([JVET-V0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10704)) will be applied [3] and test results will be regenerated (in order to simplify comparison with AhG11 anchor). Results in “all-intra” and “low-delay” configurations will remain the same as reported in JVET-V Exploration experiment.

|  |  |  |
| --- | --- | --- |
| Proposal | What to study? | SW owner |
| [JVET-V0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10798) | EE1-1.1: neural network based in-loop filter using depthwise separable convolution and regular convolution (JVET-U0061) . | Xiaozhongxu  [xiaozhongxu@tencent.com](mailto:xiaozhongxu@tencent.com) |
| [JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588) | EE: SSIM based CNN model for in-loop filtering | Tong Ouyang  [oyjiyu@whu.edu.cn](mailto:oyjiyu@whu.edu.cn) |
| [JVET-V0115](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10764) | EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer | [H. Wang](mailto:hongtaow@qti.qualcomm.com)  [hongtaow@qti.qualcomm.com](mailto:hongtaow@qti.qualcomm.com) |

To simplify comparison, structures of neural networks for all proposals in this category are shown on Fig. 1-3 below.

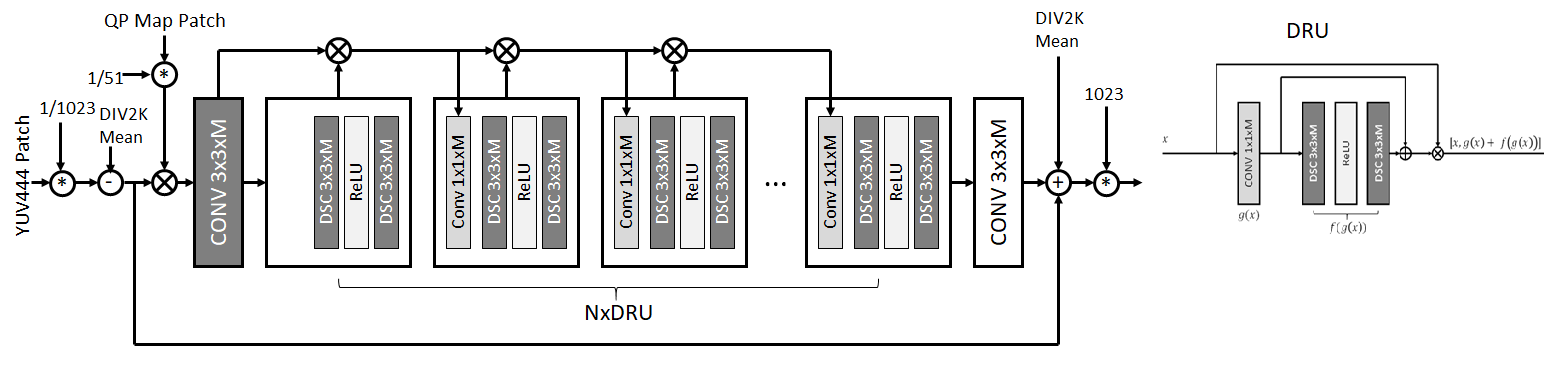


Fig. 1. Low-complexity NN-based in-loop filter with DSC from [JVET-V0137](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10798).





Fig. 2. SSIM optimized NN-based in-loop filter with SE block from [JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588).



Fig. 3. "NN-based de-block" with long activation function from JVET-V0115.

## NN-based super-resolution

The majority of proponents in this category decided not to change filter design compare to JVET-V-EE1. Only code modification related to new MCTF ([JVET-V0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10704)) [3] will be applied and test results will be regenerated (in order to simplify comparison with AhG11 anchor). Results in “all-intra” and :low-delay” configurations will remain the same as reported in JVET-V Exploration experiment.

Anchor for tests in this category will be VTM11.0 with new MCTF ([JVET-V0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10704)) with QP =27, 32, 37, 42, 47. Additionally two reference configurations (which use VTM RPR2 functionality) will be included to EE testing. The goal is to assess benefit of NN-based super resolution compared to functionality already available in VVC standard.

It is recommend to use scaling factor 2 for all experiments in this category (--ScalingRatioHor=2.0 --ScalingRatioVer=2.0).

Quality metrics must be computed in full resolution (“--UpscaledOutput=2”).

Two different configurations of RPR in VVC (depicted in Fig. 4) will be tested as reference point for NN-based super-resolution.

Additionally selection (per sequences and rate point) between two RPR cofngiuraitons will be reported, just to estimate potential of RPR as coding tool.

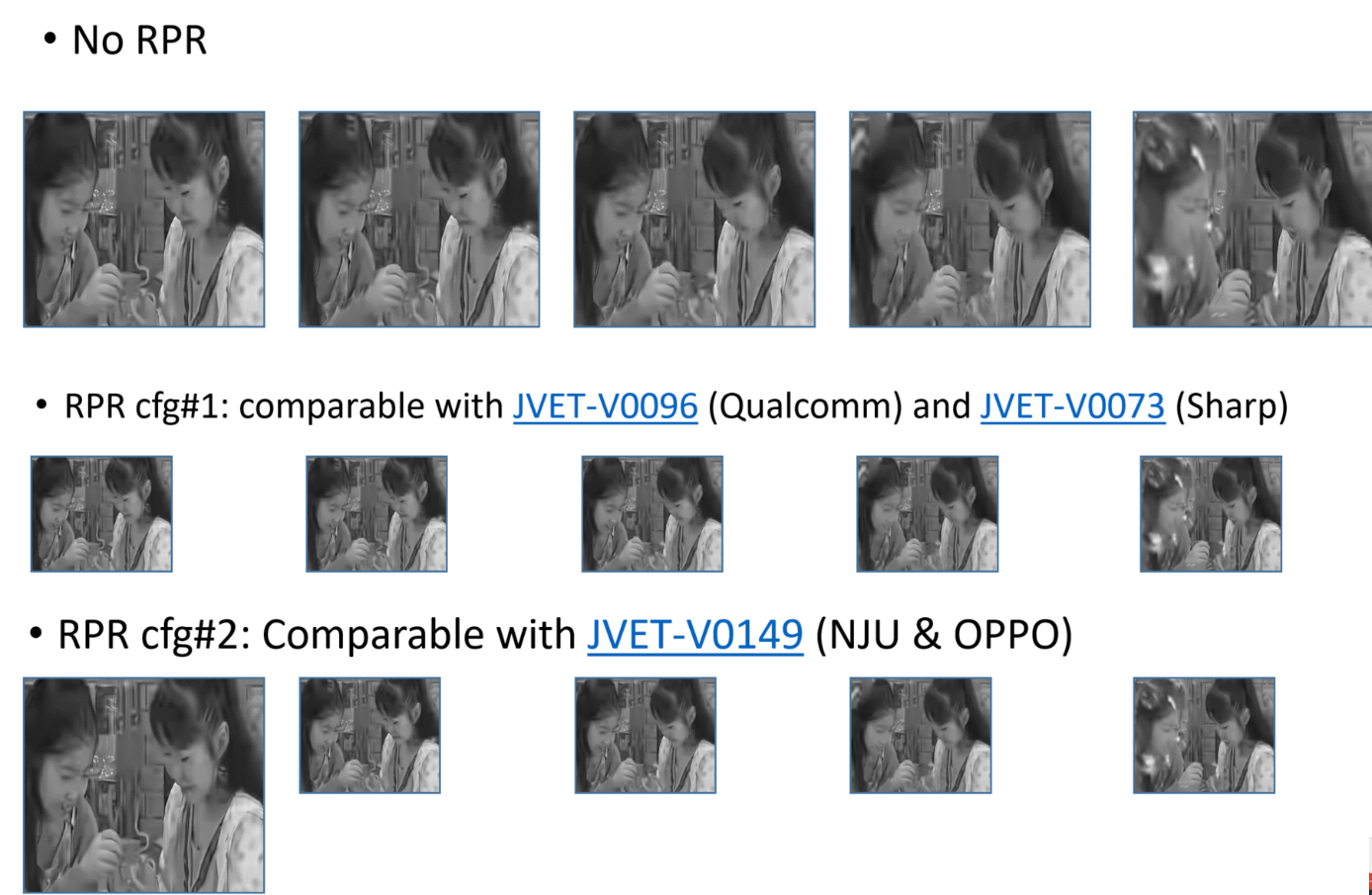


Fig. 4. RPR configurations comparable with EE tests.

|  |  |  |
| --- | --- | --- |
| Proposal(s) | What to study? | Contact |
| VVC RPR2 cfg#1 | **All frames** coded in quarter size . This configuration is natural comparsion point for [JVET-V0073](https://jvet-experts.org/doc_end_user/current_document.php?id=10721) and [JVET-V0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10744) | AhG11 |
| VVC RPR2 cfg#2 | I-frames are coded in full resolution, Inter-frames **frame** are coded in quarter size. This configuration is natural comparsion point for [JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) | AhG11 |
| Selective RPR | Selection between full size and RPR cfg#1 or cfg#2. This configuration is just for information, to check potential of RPR as coding tool. | AhG11 |
| [JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) | EE1: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-U0096) | [M. Lu](mailto:luming@smail.nju.edu.cn)  [luming@smail.nju.edu.cn](mailto:luming@smail.nju.edu.cn) |
| [JVET-V0073](https://jvet-experts.org/doc_end_user/current_document.php?id=10721) | EE1.2: Additional experimental results of NN-based super resolution (JVET-U0053) | [T. Chujoh](mailto:chujoh.takeshi@sharp.co.jp)  [chujoh.takeshi@sharp.co.jp](mailto:chujoh.takeshi@sharp.co.jp) |
| [JVET-V0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10744) | EE1-2.3: Neural Network-based Super Resolution | A. [M. Kotra](mailto:akotra@qti.qualcomm.com)  [akotra@qti.qualcomm.com](mailto:akotra@qti.qualcomm.com) |

# Visual test

Proponents will provide suggestions which video sequences to be included to the viewing. Proponents are suggested not to propose UHD content coded in LDP, LDB configuration for viewing (due to the short length, 3 sec).

* select matching rate points for anchor
* prepare and upload mp4 recommended for viewing,

rate difference with Anchor <10%.Recommendation on the YUV file upload are:

* Data is to be uploaded to the JVET ftp site using the jvet-ul1 login to the directory JVET-W\_EE-DNN (if participants need credentials, please contact Mathias Wien)
* Inside the directory, proponents can create a separate sub-directory for their proposal
* Naming convention for the YUV files: <CTC sequence file name including resolution and frame rate>\_<jvet doc number>\_<sub-EE identifier (if applicable)>\_QP<qp-value>.yuv
* YUV files shall be zipped.
* Providing md5sums for the YUV files together with the zips may help others to verify the integrity of the uploaded file

It is agreed to have AhG11 telco for discussion of the final list of sequences for viewing selection. Link to the telco is added to JVET calendar (https://jvet-experts.org/).

Total number of test points per viewing session is limited to 20. 1080p and 4K should be in different viewing sessions. In order to increase number of viewers it is suggested to crop 4K sequences to 1080p size for viewing only (cropping to be discussed during telco).

# Software and communication channel

Software for each test will be made available for all JVET members (under MPEG or VCEG password) according to deadline specified in section 6. Proponent are asked to provide short and clear description about software usage, including package needed to be installed. This SW description file shall be uploaded together with SW.

EE SW location is <https://vcgit.hhi.fraunhofer.de/jvet-w-ee1>. SW branches contain proposals number EE-JVET-V0XXX.

One SW branch will be created for each contribution included into EE. If multiple tests need to be conducted then it is recommended to use single SW branch with different command line options. Configuration of SW for each tests needs to be clearly defined in SW description file.

If different configuration use different sets of models then it also shall be described in SW description file.

After SW is uploaded and ready for review it is supposed to be announced in JVET reflector. SW modification (including models up-date after retraining) after SW availability announcement is allowed but suggested to be minimized and all up-dates have to be announced in JVET reflector again with short description of what exactly was changed and why.

If additional SW branch for combination of multiple tools in EE for maximum potential gain demonstration is needed then it can be created and combination test announced in JVET reflector.

# Timeline

**T1 = 2 weeks after JVET-V meeting (14-May-2021):** To revise EE description. Changes should be discussed and agreed on JVET reflector (should we have a telco?).

**T2 – 3 weeks after JVET-V meeting (21-May-2021)**: Anchor is available.

**T3 = 3 weeks before T5 (14-June-2021):** SW is published and frozen, technology description is ready, and cross-check starts.

**T4= 2 weeks before T5 (21-June-2021):** AhG11 telco for viewing preparation.

**T5 = 05-July-2021:** EE summary is uploaded as input contribution.

# References

[1] **Common Test Conditions and evaluation procedures for neural network-based video coding technology**, JVET-V2016.

[2] <https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/VVCSoftware_VTM>

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

[4] https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc