 ISO/IEC JTC 1/SC 29/WG 5 N0040

**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 technology

**Status:** Approved

**Date of document:** 2021-01-15

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

**Expected action:** Info

**Action due date:** None

**No. of pages: 8** (without 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**  21st Meeting, by teleconference, 6–15 Jan. 2021 | Document: JVET-U2023\_r2 |

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

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

# Abstract

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

# Introduction

In the 21st JVET meeting it was decided [1] to setup an exploration experiment on NN-based video coding and include technologies listed in section 2. Additionally to getting better understanding of each technology EE should determine the potential coding efficiency gain that is possible without regard to complexity. Combination of elements from different EE tests can be tested to accomplish this goal.

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 [2], use results reported template and fill out complexity assessment page (at least mandatory elements as requested in [2]). Namely VTM11.0 should be used as anchor, GOPsize in RA is 32, MCTF is enabled.

Cross-check of EE test is highly encouraged, since beside of results verification the availability of cross-check indicates the interest from the group to tested technology. Cross-checker is expected to study SW, comment about accuracy of technology description and complexity assessment numbers in Results Reporting Template [2]. Since AhG11 will try to create “secret test set” after EE SW freeze, the trial run using a “secret test set” supposed to the part of a cross-check.

Proponents must are encouraged to test technology using materials out of NN-CTC [2] and report their observation to the group. Cross-checker is expected to study SW, comment about accuracy of technology description and complexity assessment numbers in Results Reporting Template [2].

Detailed information about training beyond mandatory in [2] (such as training sub-set selection logic, learning curve) is encouraged to be provided by proponent.

Proponents are highly encouraged to respects and models freeze deadline, in order make AhG11 possible to exercise “secret test set”.

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

# List of experiments

## NN-based filtering

Proponents in this category are mandated to report if MCTF was enabled during training; and invited to provide results w/o MCFT.

|  |  |  |
| --- | --- | --- |
| Proposal | What to study? | SW owner |
| [JVET-U0061](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10575&c=E,1,fautfX8-Mw5qWF1uMHMdxkvhGYv-3ofTckxdlyE9olsgneDdpIEvIwiN-XiBzMd92kmn-djxHwIEUrD6W4bWw78SedlhMeUrqtZTOSB8qg98Q2siPg,,&typo=1) | * Effect of clipping as introduced in [JVET-U0061](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10575&c=E,1,fautfX8-Mw5qWF1uMHMdxkvhGYv-3ofTckxdlyE9olsgneDdpIEvIwiN-XiBzMd92kmn-djxHwIEUrD6W4bWw78SedlhMeUrqtZTOSB8qg98Q2siPg,,&typo=1). | [C. Auyeung](mailto:cauyeung@tencent.com) |
| [JVET-U0074](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10588&c=E,1,jpTDckiZevVC2CrkJoA3cF1uGGPLeenjIz9cccXp6Rx_AFdD40Dd__tfJgXPNK_ayht_Zi1lXOZhv2fhahWsfd9zAbNPBgE7uUbpObp1Stu5A0ovlA,,&typo=1) |         Model trained with L1 loss          Model trained with loss function includes SSIM | [T. Ouyang](mailto:oyjiyu@whu.edu.cn) |
| [JVET-U0115](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10642&c=E,1,1T80n_p5VuTSSJZjv1NdXJzBqcn82-zm1DZDnqu6P0tjKpwlu_IuAV-vCzrNUUVnFHG6vM6u5wpu48Gl9VbtdIiFTkPYKK2-2CwSU5sX01LWpQaR&typo=1) | * Different operation points (complexity/performance) | [H. Wang](mailto:hongtaow@qti.qualcomm.com) |
| [JVET-U0104](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10619&c=E,1,ouq0br_KDMzlC2JcnvdhMUijEwcTInTvZXKNWzMoQ1wcNckbcZneBeNJP8qJgygwVz1imY8e-vh95sO1fPVSl0yl-IvWqA6W7Q0PuPBS9fNvOCdlX7t_HWNq6vs,&typo=1) | * Different operation points (complexity/performance) | [H. Wang](mailto:hongtaow@qti.qualcomm.com) |

[JVET-U0061](https://jvet-experts.org/doc_end_user/current_document.php?id=10575) EE-1.1-related: BD-Rate improvements to JVET-T0057 neural network based in-loop filter using depthwise separable convolution and regular convolution [C. Auyeung, X. Li, S. Liu (Tencent)]

Algorithm description:

* Depth-wise separable convolution, 4 residual blocks, up-sampled Chroma
* After up-sampling values outside of the range are clipped out

Suggested tests:

Test 1. With clipping (as code as JVET-U0061, but tested under updated NN CTC)

Cross-checker(s):

[JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588) EE: SSIM based CNN model for in-loop filtering [T. Ouyang, H. Zhu, Z. Chen (Wuhan Univ.), X. Xu, S. Liu (Tencent)]

Algorithm description:

* Depth-wise separable convolution, 8 residual blocks, SE block, up-sampled Chroma
* Normalized QP map is additional (to color planes) input to NN
* During training loss-function is weighted sum of SSIM and L1 distortion measure
* 

Suggested tests:

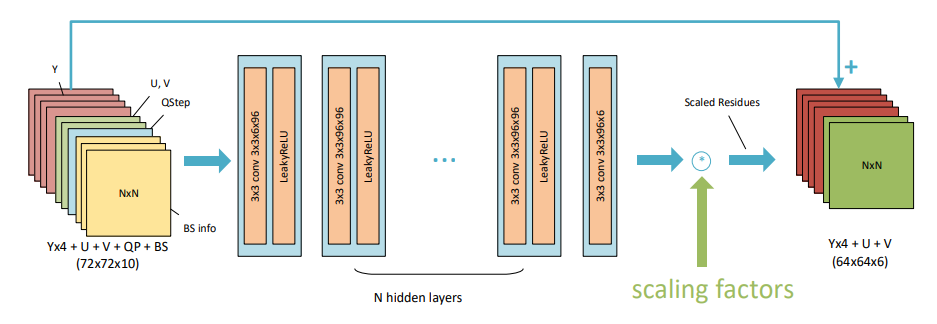
Test 1: Best performing version (code based on JVET-U0074, but tested under updated NN CTC, MS-SSIM performance reported)

Visual quality investigation is recommended.

[JVET-U0115](https://jvet-experts.org/doc_end_user/current_document.php?id=10642) AHG11: Neural Network-based In-Loop Filter Performance with No Deblocking Filtering stage [H. Wang, J. Chen, A. M. Kotra, M. Karczewicz (Qualcomm)] [late]

Algorithm description:

* Model process extended (+8 samples on each side) block
* Luma split to 4 planes
* BS, and QStep are additional (to color planes) NN input
* Frame level NN model selection (from 4 candidates)
* Slice and CTU on/off flag
* Scaling factor per color component selected at picture level
* De-blocking is disabled (not needed with this approach)



Suggested tests:

Test 1: Tests with different model sizes.

Test 2: CTU level model control

Visual quality investigation is recommended.

Cross-checker(s):

[JVET-U0104](https://jvet-experts.org/doc_end_user/current_document.php?id=10619) AHG11: In-loop filtering with convolutional neural network and large activation [J. Chen, H. Wang, A. M. Kotra, M. Karczewicz (Qualcomm)]

Algorithm description:

* Luma split to 4 planes
* Frame level NN model selection (from 4 candidates)
* Slice and CTU on/off flag
* Scaling factor per color component selected at picture level
* ReLU has been replaced by “long activation function” (Conv1×1×K×M, LeakyReLu, Conv1×1×M×K).

Parameters M, K and N- number of hidden layers – are subject to study performance/ complexity tradeoff.



Suggested tests:

Test 1: Tests with different model sizes

Test 2: CTU level model control.

Cross-checker(s):

## NN-based super-resolution

Anchor for tests in this category will be VTM11.0 with QP =27, 32, 37, 42, 47

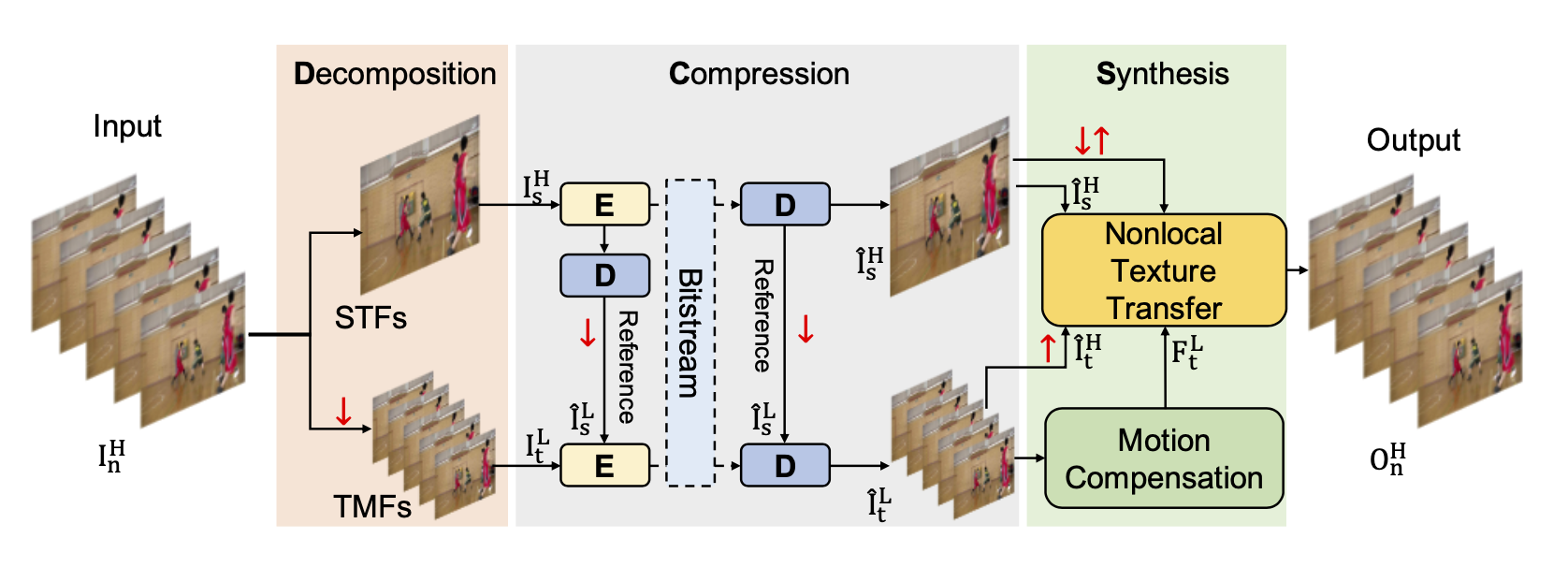
Proponents in this category are encouraged to perform comparison with RPR functionality enabled in VTM (performance of VTM11.0 with RPR ×2↓to be provided by AhG11).

|  |  |  |
| --- | --- | --- |
| Proposal(s) | What to study? | Contact |
| [JVET-U0096](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10611&c=E,1,rT0RMQ5BNUE6XgAWLg4DWWdkRE_MvN5dxIjKYrbD-9bnrEFOsSiMF5S4m2PCWikois47TGaJ0_ysPt-xuaT4yDUVAakAP2hRmteQ7pwHdIRp&typo=1) | * Chroma performance to be reported * Extension to RA configuration * NN scheme adjustment in presence of MCTF | [M. Lu](mailto:luming@smail.nju.edu.cn) |
| [JVET-U0053](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10565&c=E,1,dvpzDj61QnY378GhCwtFwU_a7U78V1ChhIC9jUg98Zk315gDf0RvnNIQkVnaJo-Y1Ho9ZPOVPUjx-ncCbtbT6PvFcvbPfFj9Zp7GXDgcPhw,&typo=1) | * Sending network parameters, update the network by sequences. | [T. Chujoh](mailto:chujoh.takeshi@sharp.co.jp) |
| [JVET-U0099](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fjvet-experts.org%2fdoc_end_user%2fcurrent_document.php%3fid%3d10614&c=E,1,W-jIoncKO7nBzB-NmpcwswUiyI1ziNut-s54f9aeQpgpM9-1ciBzA_RVFfAq5gUzlcUVaLrlviKBeLbDd3o3beLpSiRautamSl-OBdfLx4gGfaiL9iUiYEFAlQ,,&typo=1) | * Sequence based super resolution (code based on JVET-U0099, but tested under updated NN CTC) * Picture based RD selection of either downsampled resolution (or) normal resolution * CTU based RD selection of either downsampled resolution (or) normal resolution | A. [M. Kotra](mailto:akotra@qti.qualcomm.com)  K. Reuzé |

[JVET-U0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10611) EE: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology [M. Lu, Z. Ma (Nanjing Univ.), L. Xu, D. Wang (OPPO)]

Algorithm description:

* Base-line codec agnostic
* I-frame coded in full resolution
* Temporally predicted frames are coded in quarter size
* Full-size video synthesized with
  + Multi-frame motion compensation
  + Non-local texture Transfer



Suggested tests:

Test 1: Best performing version (code based on JVET-U0096, but tested under updated NN CTC)

Visual quality investigation is recommended.

Cross-checker(s):

[JVET-U0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10565) AHG9/AHG11: Level information for super-resolution neural network [T. Chujoh, E. Sasaki, T. Suzuki, T. Ikai (Sharp)]

Algorithm description:

This contribution is based on a framework for introducing a super-resolution post-filter using CNN (Neural Network). In VVC, RPR (Reference Picture Re-sampling) has been introduced, and in several 4K sequences with low bit-rate, there are some coding gains by changing the resolution of the whole sequence. In this experiment, instead of the up-sampling filter, a super-resolution network is used, and it is tested that sending network parameters, update the network by sequences.



Suggested tests:

Test 1: Best performing version (code based on JVET-U0053, but tested under updated NN CTC) with NN parameters same for all videos

Test 2: NN parameters are updated per video and signalled to decoder in bit-stream

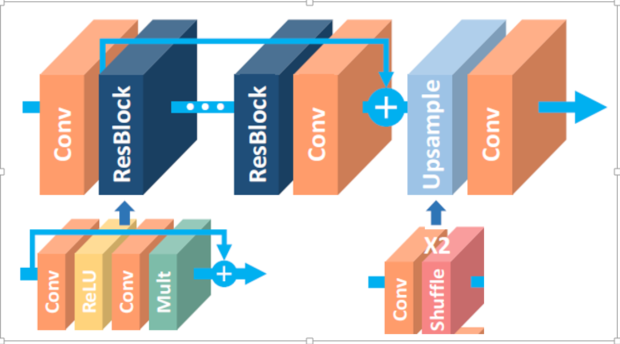
Visual quality investigation is recommended.

Cross-checker(s):

[JVET-U0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10614) AHG11: Neural Network-based Super Resolution [A. M. Kotra, K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm)]

Algorithm description:

* Conceptually the same as RPR in VVC, instead standardized predetermined coefficient up-sampling filter in RPR learnable super-resolution NN is used
* YUV4:2:0 converted to YUV 4:4:4



Suggested tests

Test 1: Sequence based super resolution (code based on JVET-U0099, but tested under updated NN CTC)

Test 2: Picture based RD selection of either downsampled resolution (or) normal resolution

Test 3: CTU based RD selection of either downsampled resolution (or) normal resolution

Cross-checker(s): Sharp ([T. Chujoh](mailto:chujoh.takeshi@sharp.co.jp))

# Visual test

Proponents provide suggestions which video sequences to be included to the viewing:

* upload yuv 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-V\_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 identity of the uploaded file

It is agreed to have AhG11 telco for discussion 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.

# 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 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-u-ee1>. SW branches contain proposals number EE-JVET-U0XXX.

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 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-U meeting (29-January-2021):** To revise EE description and refine questions to be answered. Questions should be discussed and agreed on JVET reflector and telco at 22-January-2021.

**T1.1 – 3 weeks after JVET-U meeting**: Anchor is also available.

**T2 = 4 weeks after T1 (26-February-2021):** First version of SW is available and announced in JVET reflector, and cross-checkers start to install SW.

**T3 = 3 weeks after T2 (19-March-2021):** SW is frozen, technology description is ready, and cross-check starts.

**T4= 1 weeks before T5 (06-April-2021):** AhG11 telco for viewing preparation.

**T5 = 13-April-2021:** EE summary is uploaded as input contribution.

# References

[1] **BoG Report: EE for Neural Networks**, A. Segall, JVET-U0141.

[2] **Common Test Conditions and evaluation procedures for neural network-based video coding technology**, JVET-U2016.