|  |
| --- |
| **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 148** |
| **Online, 13–22 July 2022** |
| |  |  | | --- | --- | | **Title:** | **Exploration experiment on neural network-based video coding (EE1)** | | **Source:** | **Convenor (Jens-Rainer Ohm)** | | **Type:** | **General** | | **Subtype:** | **N/A** | | **Status:** | **Approved** | | **Date:** | **2022-08-06** | | **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**  27th Meeting, by teleconference, 13–22 July 2022 | Document: JVET-AA2023-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Exploration Experiment on Neural Network-based Video Coding (EE1)** | | |
| *Status:* | Output Document to JVET | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | E. Alshina, F. Galpin, Y. Li, M. Santamaria, H. Wang, L. Wang and Z. Xie | Tel: Email: | [elena.alshina@huawei.com](about:blank)  [Franck.Galpin@InterDigital.com](mailto:Franck.Galpin@InterDigital.com)  [yue.li@bytedance.com](mailto:yue.li@bytedance.com)  [maria.santamaria\_gomez@nokia.com](mailto:maria.santamaria_gomez@nokia.com)  [hongtaow@qti.qualcomm.com](about:blank)  [liqiangwang@tencent.com](mailto:liqiangwang@tencent.com)  [xiezhihuang@oppo.com](mailto:xiezhihuang@oppo.com) |
| *Source:* | EE coordinators | | |

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

# Abstract

This document describes Exploration Experiments (EEs) planned to be performed between the JVET-AA and JVET-AB 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-AA meeting.

NNVC common SW base (**NCS**) and training scripts are encouraged to all AhG11 and EE1 proponents, but not mandated for this round of EE1 due to the limited time.

The most promising technologies recommended by JVET **will undergo procedure of cross-check for the training**.

Tests will be conducted in three categories: enhancement filters (both post- and in-loop), super-resolution and E2E NN-based video coding.

All proponents **must** report results relatively to AhG11 anchor <https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/VVCSoftware_VTM/-/tags/VTM-11.0_nnvc-2.0> [2] (VTM-11.0 + “newMCTF” + “De-blocking filter RDO”), QP=22, 27, 32, 37, 42) and the reported template recommended by AhG11.

If test is improving or modifying JVET-AA0088 or / and JVET-AA0111 NN-filters, included into NCS, then results relatively to JVET-AA0088 or / and JVET-AA0111 (as appropriate) are requested.

Proponents in Super Resolution category are requested to report results both relatively to AhG11 anchor, but also relatively to adaptive resolution coding with RPR filter [6].

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

In this category 4 tests are based on filter architecture JVET-AA0111 (included into NCS). One test is based on JVET-AA0088 (included into NCS). Another 3 tests use conceptually different architectures from those in NCS. Three tests supposed to have **training cross-check**.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Test | Proposal | Proponent | Anchor | Training xCheck | Inference | Contact(s) | Cross-checker |
| 1.1 | [JVET-AA0066](https://jvet-experts.org/doc_end_user/current_document.php?id=11742) | Nokia | NNVC | YES | SADL | Maria Santamaria |  |
| 1.2 | [JVET-AA0081](https://jvet-experts.org/doc_end_user/current_document.php?id=11757) | Ericsson | NNVC | YES | SADL | [Jacob Ström](mailto:jacob.strom@ericsson.com) |  |
| 1.3 | [JVET-AA0074](https://jvet-experts.org/doc_end_user/current_document.php?id=11750) | Xidian, OPPO | NNVC |  |  | [H. Zhang](mailto:13227706628@163.com) | Maria Santamaria |
| 1.4 | [JVET-AA0112](https://jvet-experts.org/doc_end_user/current_document.php?id=11788) | Bytedance, | NNVC and AA0111 |  | SADL | [Yue Li](mailto:yue.li@bytedance.com) |  |
| 1.5 | [JVET-AA0090](https://jvet-experts.org/doc_end_user/current_document.php?id=11766) | Ericsson | NNVC and AA0111 |  | SADL | Du Liu |  |
| 1.6 | [JVET-AA0113](https://jvet-experts.org/doc_end_user/documents/27_Teleconference/wg11/JVET-AA0113-v2.zip) | Bytedance, | NNVC and AA0111 | YES | SADL | [Junru Li](mailto:lijunru@bytedance.com) |  |
| 1.7 | [JVET-AA0131](https://jvet-experts.org/doc_end_user/documents/27_Teleconference/wg11/JVET-AA0131-v3.zip) | Qualcomm | NNVC and AA0111 |  | SADL | [Samuel Eadie](mailto:seadie@qti.qualcomm.com)/  [Hongtao Wang](mailto:hongtaow@qti.qualcomm.com) |  |
| 1.8 | [JVET-AA0089](https://jvet-experts.org/doc_end_user/current_document.php?id=11765) | Tencent | NNVC and AA0088 |  | SADL | [Liqiang Wang](mailto:liqiangwang@tencent.com) |  |

**Test JVET-AA-EE1-1.1** [**JVET-AA0066**](https://jvet-experts.org/doc_end_user/current_document.php?id=11742) **EE1-1.7: Content-adaptive post-filter based on SADL inference [R. Yang, M. Santamaria, F. Cricri, H. Zhang, J. Lainema, R. G. Youvalari, M. M. Hannuksela (Nokia)]**

This test studies the NN post-filter proposed in JVET-AA0066. The inputs to the filter are the reconstructed luma samples, chroma samples and frame QP.



Aspects to study:

Test 1.1.1: Study the quantisation of the post-filter in JVET-AA0066. Additionally, the signalling is done using only the post-filter SEI messages (characteristics and activation), compliant with JVET-Z0244, which means the ON/OFF flags (CTU and slice level) and scaling factors (picture header) from JVET-AA0066 are discarded.

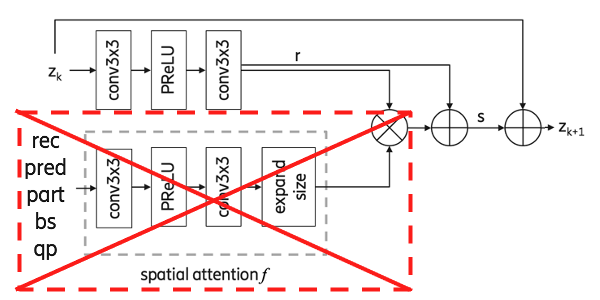
**Test JVET-AA-EE1-1.2 JVET-AA0081 EE1-1.2: NN intra model without attention, partitioning and boundary strength.**

This test evaluates whether the attention branch and partitioning input can be removed from the intra luma model of JVET-Z-EE1-1.6 (JVET-AA0111) with preserved performance. The first change is to remove the prediction in put according to the figure below.

Diagram

Description automatically generated

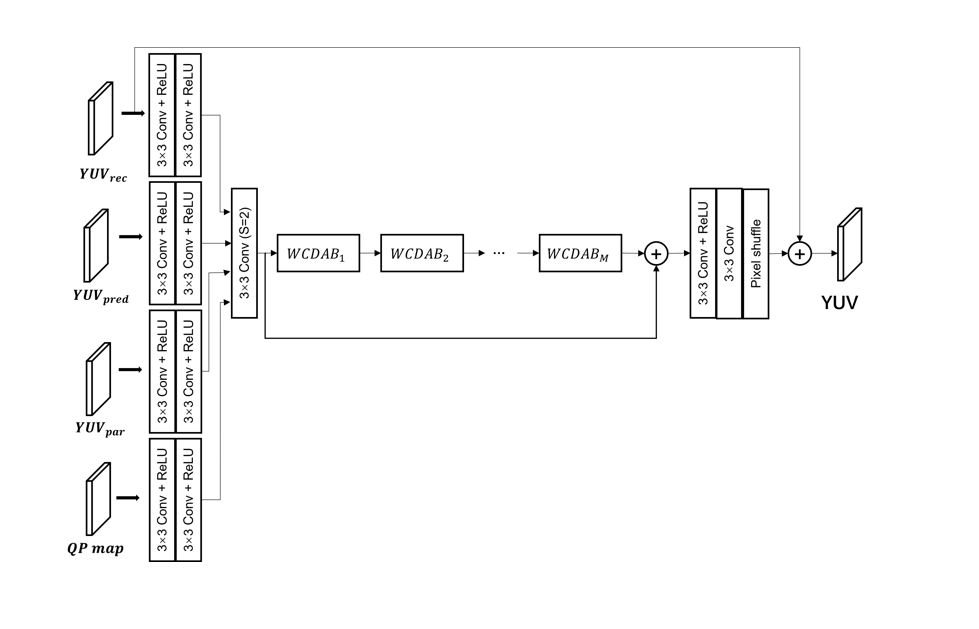
The second change is to remove the attention branch according to the figure below.

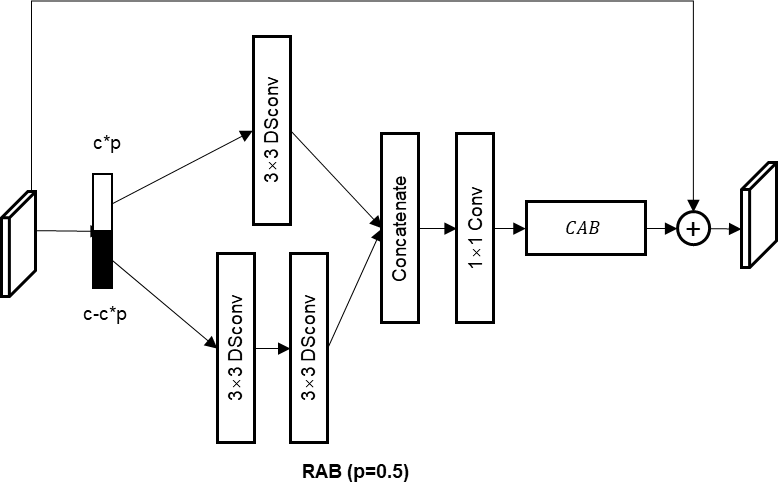


Tests to be studied in EE:

Test 1.2.1: Verify training of the model from JVET-Z-EE1-1.2.1, i.e., with no prediction input in for the intra luma model and with no attention branch for the intra luma model. The inference is going to be done in SADL.

**Test JVET-AA-EE1-1.3 JVET-AA0074 [EE1-related] Lighter WCDANN: CNN Based In-Loop Filters**





**Tests to be studied in EE:**

**Test 1.3.1 Test the network architecture as proposed in JVET-AA0074 (AI configuration).**

**Test 1.3.2 Test the network architecture by varying weights in loss function (AI configuration).**

**Test JVET-AA-EE1-1.4** [**JVET-AA0112**](https://jvet-experts.org/doc_end_user/current_document.php?id=11788) **EE1-1.6-related: Deep In-Loop Filter with Additional Input Information [Y. Li, K. Zhang, L. Zhang (Bytedance)]**

This test investigates the neural network-based in-loop filtering method proposed in JVET-AA0112, which is developed on top of [4]. Two aspects are proposed to improve the CNN models presented in [4]. Aspect #1: Addition input information can be input to the in-loop filtering network. Aspect #2: The input samples of CNN can be flipped and the output samples of CNN can be flipped back.The network backbone is based on residual blocks. Other design elements including parameter selection, residual scaling, combination of NN filtering and deblocking filtering, etc. remain the same as [4].



Tests to be studied in EE

Test 1.4.1: Test the networks in JVET-AA0112 on top of the common software base.

Test 1.4.2: Optimize performance-complexity tradeoff (e.g. number of res blocks, input information, ResBlock structure, number of models, etc.)

**Test JVET-AA-EE1-1.5** [**JVET-AA0090**](https://jvet-experts.org/doc_end_user/current_document.php?id=11766) **EE1-related: One luma model with IPB and skip for filtering intra and inter luma slices [D. Liu, J. Ström, M. Damghanian, P. Wennersten, Y. Li (Ericsson)]**

This test evaluates the luma model which uses one luma model with block type information and block skip information for NN-filtering of intra luma and inter luma slices. The proposed model uses block type information IPB (intra/uni-predicted/bi-predicted) and block skip information (bypassed or not) as two additional inputs on top of the inter luma model from JVET-Z0112.

Diagram

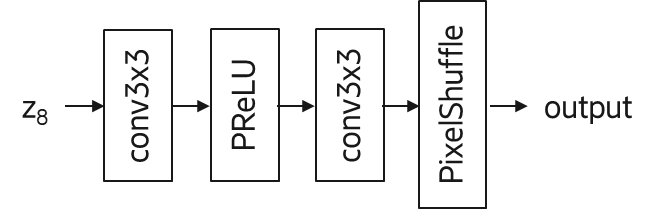
Description automatically generated

*Figure 1. Head of network. The inputs are combined to form the input y to the next part of the network.*

Diagram

Description automatically generated

*Figure 2. The k-th residual block (k=0..7). The output y of the head is fed into a first residual block with input z0=y, which also takes the inputs rec, pred, part, bs, qp, IPB, and skip. The output z1 is then fed into another such residual block.*

**  
*Figure 3. The output of the last residual block is fed into this last part of the network.*

Aspects to study:

Test 1.5.1: Test the IPB+skip model used for intra luma and inter luma slices as proposed in JVET-AA0090.

Test 1.5.2: Test the IPB model (Test 1.5.1 model without skip) used for intra luma and inter luma slices.

Test 1.5.3: Test the skip model (Test 1.5.1 model without IPB) used for intra luma and inter luma slices.

**Test JVET-AA-EE1-1.6** [**JVET-AA0113**](https://jvet-experts.org/doc_end_user/current_document.php?id=11789) **EE1-1.6-related: RDO Considering Deep In-Loop Filter with SADL [J. Li, Y. Li, K. Zhang, L. Zhang (Bytedance)]**

This test evaluates encoder optimization technique on top of the deep learning-based in-loop filtering method in JVET-AA0113. A simplified version of CNN models is additionally trained and used in the RD stage where the simplified models are implemented with SADL using fixed point-based calculation.

Tests to be studied in EE:

Test 1.6.1: Test the simplified models used for RD stage as proposed in JVET-AA0113.

Test 1.6.2: Test the simplified models used for RD stage where the network structure is improved.

Test 1.6.3: Test the simplified models with different model number.

Test 1.6.4: Test the combination of Test 1.6.2 and Test 1.6.3.

**Test JVET-AA-EE1-1.7** [**JVET-AA0131**](https://jvet-experts.org/doc_end_user/current_document.php?id=11807) **EE1-related: CNN based in-loop filtering with large activation layer [H. Wang, S. Eadie, M. Coban, M. Karczewicz (Qualcomm)]**

This test evaluates the NN based loop filter proposed in JVET-AA0131, which uses the method studied in [4] is used as a baseline. The network backbone is replaced by residual blocks with large activation layers [5] as shown in the following figure. The QP value used for the inference of Cb, Cr components are kept the same. And residual scaling is performed to combine the output of deblocking filter and NN filter.



Tests to be studied in EE

Test 1.7.1: Test the same network structure as that in JVET-AA0131.

Test 1.7.2: Optimize the trade-off between performance and complexity by modifying the network structure (e.g. number of res blocks, input information, ResBlock structure etc).

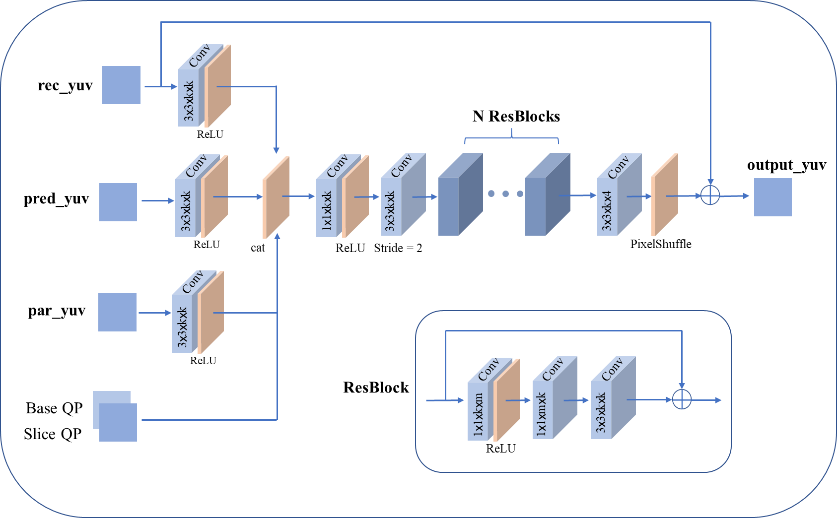
Test 1.7.3: Optimize the trade-off between performance and complexity by reducing the total number of models

Test 1.7.4: A combination of Test 1.7.2 and Test 1.7.3

Test 1.7.5: Optimize the floating-point/fixed-point SADL implementation of the models.

**Test JVET-AA-EE1-1.8** [**JVET-AA0089**](https://jvet-experts.org/doc_end_user/current_document.php?id=11807) **EE1-related: More refinements on EE1-1.4 and EE1-1.5 [L. Wang, S. Lin, X. Xu, S. Liu (Tencent), Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]**

More refinements are presented to improve the trade-off of both two-model filter and one-model filter [5]. A slice-level QP-adjustment method is used to further optimize the trade-off. Multiple results are inferred by three candidates for the B slice, and these models are indexed by {BaseQP, BaseQP - 5, BaseQP + 5}, where the index is decided at the slice-level. The current network framework is provided as the following figure.



Neural network structure of JVET-AA0089

Aspects to study:

Test 1.8.1: Test the same solution as in JVET-AA0089 on the common base software.

Test 1.8.2: Optimize the trade-off of JVET-AA0089 by improving the implementation architecture.

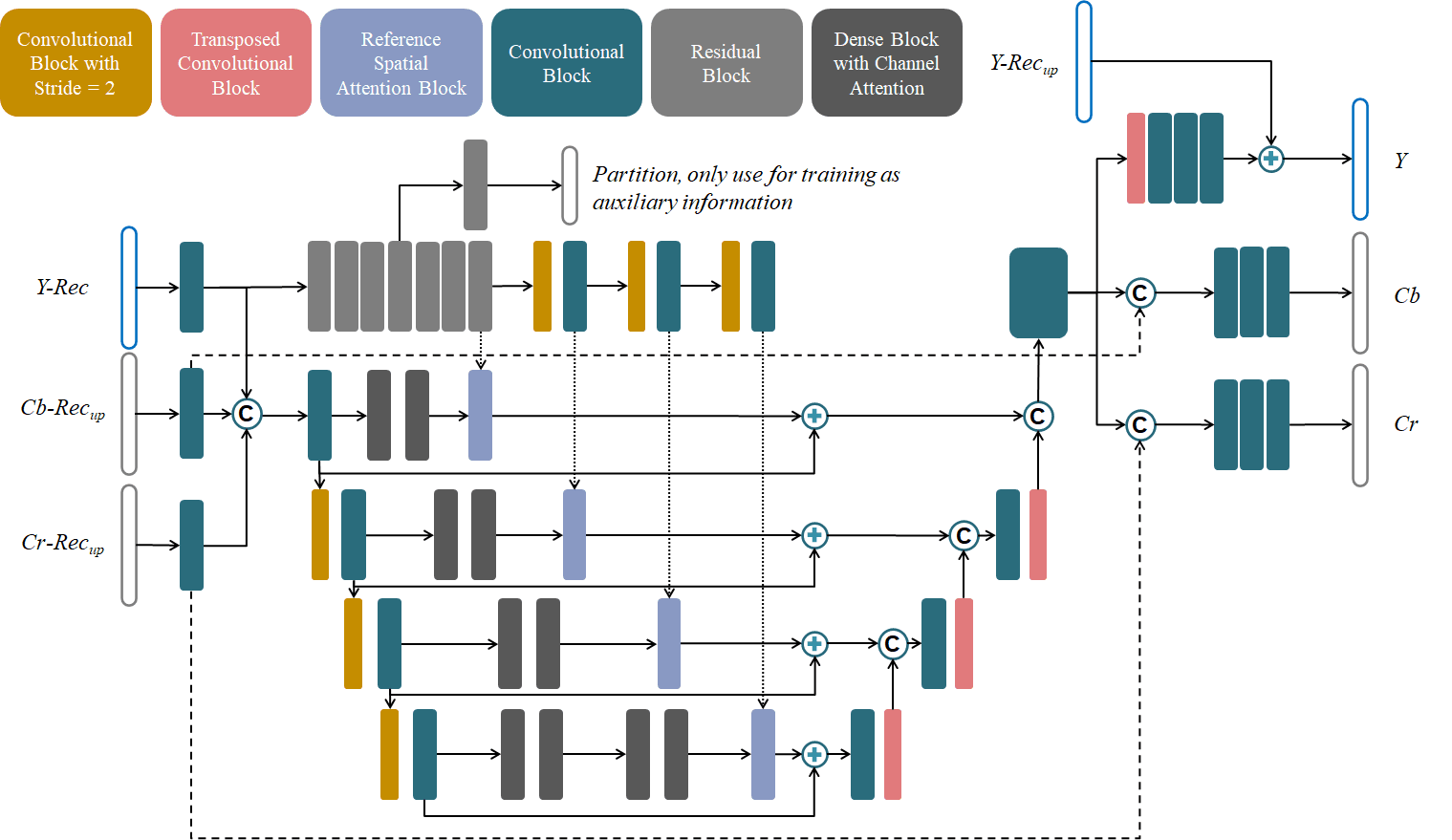
Test 1.8.3: Optimize the trade-off of JVET-AA0089 by improving the network design.

# Exploration experiments on NN-based super- resolution technology

It was requested by the group to NN-based super-resolution on top of adaptive resolution coding demonstrated in EE1 earlier and described in [6].

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test | Proposal | Proponent | Anchor | Contact | Cross-checker |
| 2.1 | [JVET-AA0076](https://jvet-experts.org/doc_end_user/current_document.php?id=11752) | Xidian Uni, OPPO | NNVC / JVET-Z0065 | [Q. Han](mailto:hanqihui2013@163.com)/ [J. Nam](mailto:junghak.nam@lge.com) |  |
| 2.2 | JVET-AA0065 | Xidian Uni, OPPO | NNVC/ JVET-Z0065 | [Shimin Huang](mailto:hsm_359@163.com)/ [J. Nam](mailto:junghak.nam@lge.com) |  |
| 2.3 | [JVET-AA0084](https://jvet-experts.org/doc_end_user/current_document.php?id=11760) | Tencent | NNVC/ JVET-Z0065 | [R. Chang](mailto:renjiechang@tencent.com)/ [J. Nam](mailto:junghak.nam@lge.com) |  |

**Test JVET-AA-EE1-2.1 JVET-AA0076 RPR-Based Super-Resolution Guided by Partition Information**

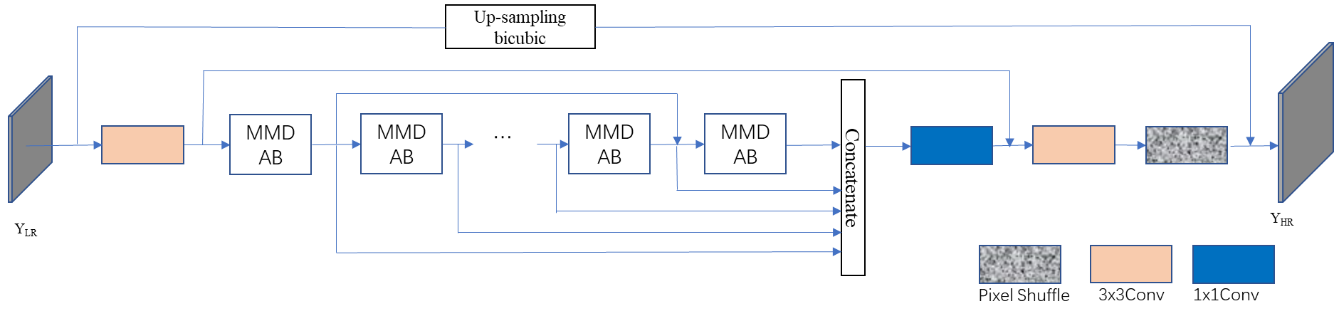


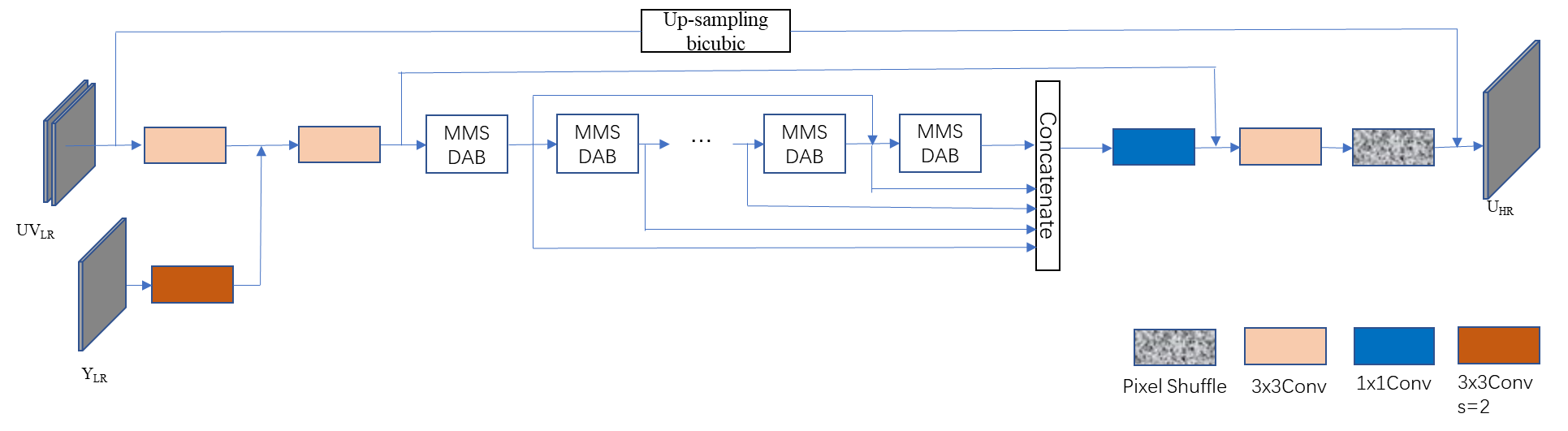
Tests to be studied in EE:

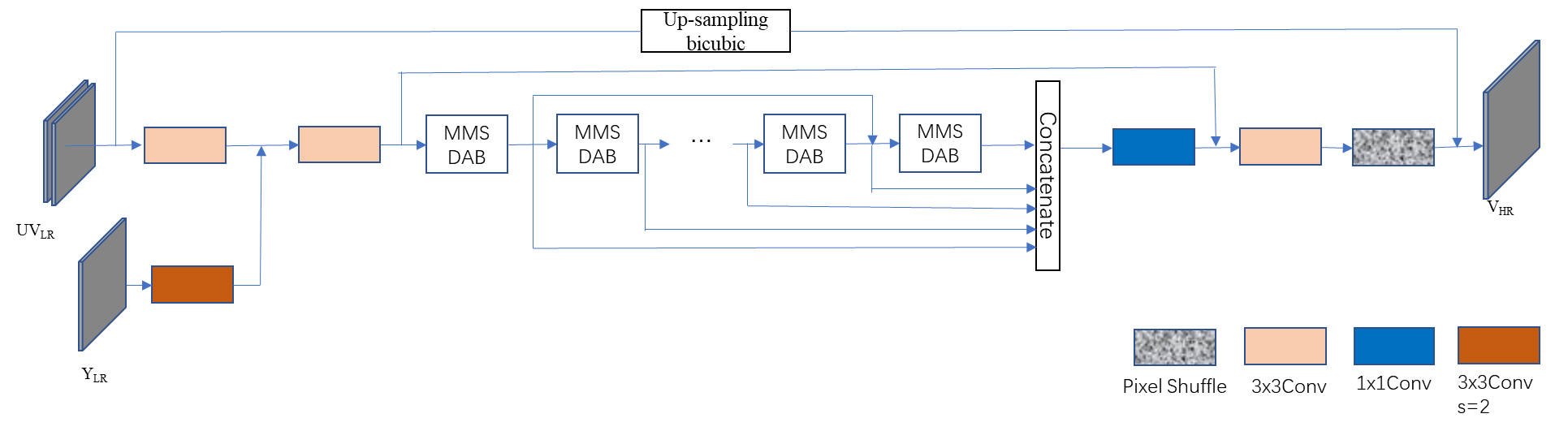
Test 2.1.1: Test the network architecture as proposed in JVET-AA0076.

Test 2.1.2: Test the network architecture without spatial attention and compare the performance.

**Test JVET-AA-EE1-2.2 JVET-AA0065 AHG11: CNN Filter for Super-Resolution with RPR functionality in VVC**







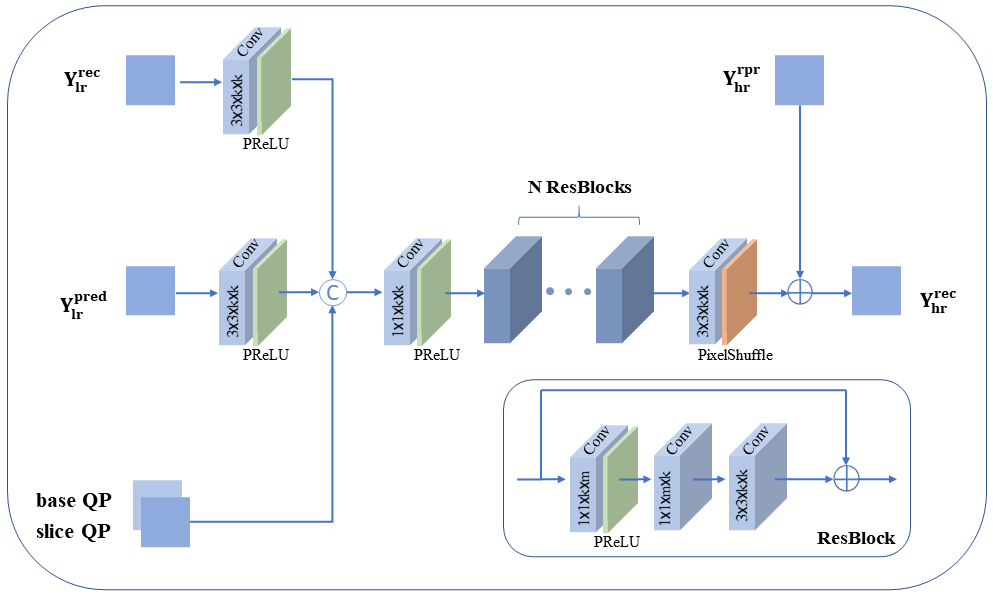
Tests to be studied in EE:

Test 2.2.1: Test the network architecture as proposed in JVET-AA0065.

Test 2.2.2: Optimize the network architecture to maintain the gain while reducing the number of parameters and complexity.

**Test JVET-AA-EE1-2.1** [**JVET-AA0084**](https://jvet-experts.org/doc_end_user/current_document.php?id=11760) **AHG11: Neural Network based Super Resolution for Video Coding Using Multiple Side Information**

This test evaluates the proposed super resolution method in JVET-AA0084. During the encoding process, the input video is first down-sampled by RPR filter and encoded. The proposed super resolution network is then used to generate the up-sampled video during the decoding process. In addition to the reconstruction, multiple side information including prediction, slice QP, base QP and slice type is also fed into the network. Besides, considering the different characteristics between luma and chroma component, different models are designed for luma and chroma component.



Neural network structure of JVET-AA0084

Aspects to study:  
Test 2.3.1: Test the combination method with RPR (JVET-Z0065) for bit-rate matching.

Test 2.3.2: Optimize the trade-off of JVET-AA0084.

# Exploration experiments on E2E NN-based video coding technology

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test | Proposal | Proponent | Anchor | Contact | Cross-checker |
| 3.1 | [[JVET-AA0059](https://jvet-experts.org/doc_end_user/current_document.php?id=11735)](https://jvet-experts.org/doc_end_user/current_document.php?id=11512) | Xidian Uni., OPPO | NNVC | [Qipu Qin](mailto:kippqin@stu.xidian.edu.cn) | Arora, Arjun [Arjun.Arora@dolby.com](mailto:Arjun.Arora@dolby.com) |

**Test JVET-AA-EE1-3.1 JVET-AA0059 Supplementary experiments based on JVET-Z0077**

Suggestion to encode key-frames as E2E AI codec or/and with VVC P or B-frames.

Test to be studied in EE：

Test 3.1.1: Test the network architecture as proposed in JVET-AA0059**.**

# 

# Visual test

No visual test is planned for this round of EE.

# Cross-check of training

Two tests in this EE (EE1-1.1, EE1-1.2 and EE1-1.6) are considered supposed to undergo cross-check of training. Training scripts provided by proponents will be used for this cross-check. Training cross-check should be described as recommended in [2].

Training scripts which allow to re-produce model parameters using training data agreed in AhG11 expected to be uploaded together with SW by T2 (refer to “Timeline” section). During one week period the cross-check of training is expected start. Proponents of EE1-1.1, EE1-1.2 and EE1-1.6 are kindly asked to provide necessary help and explanation regarding training to the cross-checkers.

# Timeline

**T1 - 2 weeks after JVET-AA meeting (05-Aug-2022):** To revise EE description. Changes should be discussed and agreed on JVET reflector. Anchor is available.

**T2 – 3 weeks after JVET-AA meeting (12-Aug -2022)**: Initial software release (which includes training scripts for proposals supposed to have training verified) that matches what was proposed to the meeting.

**T3 – 5 weeks after JVET-AA meeting (26-Aug -2022)**: Sufficient explanation about training scripts has been provided to the cross-checker and training verification starts (tests EE1-1.1, EE1-1.2 and EE1-1.6 only).

**T4 - ~4 weeks before T6 (15-Sept-2022):** Software is frozen (and may include improvements), technology description is ready, and cross-check of compression performance test starts.

**T5 – 3 days before T6 (09-Oct-2022):** Cross-checkers report status to EE1 coordinators.

**T6 - 12-Oct-2022:** EE1 summary is uploaded as input contribution.

# . References

[1] BoG Report: Neural Networks Video Coding Analysis and Planning, JVET-X0188.

[2] Common Test Conditions and evaluation procedures for neural network-based video coding technology, JVET-AA2016.

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

[4] EE1-1.6: Deep In-Loop Filter With Fixed Point Implementation, JVET-AA0111, Y. Li, K. Zhang, J. Li, L. Zhang (Bytedance), H. Wang, M. Coban, A.M. Kotra, M. Karczewicz (Qualcomm), F. Galpin (InterDigital), K. Andersson, J. Ström, D. Liu, R. Sjöberg (Ericsson)

[5] EE1-1.5: Neural network based in-loop filter with a single model, JVET-AA0088, L. Wang, S. Lin, X. Xu, S. Liu (Tencent), F. Galpin (InterDigital)

[6] [J. Nam](mailto:junghak.nam@lge.com), [S. Yoo](mailto:sunmi.yoo@lge.com), [J. Lim](mailto:jaehyun.lim@lge.com), [S. Kim (LGE)](mailto:seunghwan3.kim@lge.com) EE1-2.1: RPR encoder with multiple scale factors, [JVET-Z0065](file:///C:\Users\e00443164\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\Z6QCQSWM\JVET-Z0065) <https://vcgit.hhi.fraunhofer.de/jvet-y-ee1/VVCSoftware_VTM/-/tree/EE1-2.1>