|  |
| --- |
| **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 207** |
| **Antalya, TR – 21–28 April 2023** |
| |  |  |  | | --- | --- | --- | | **Title:** | **Exploration experiment on neural network-based video coding (EE1)** | **Insert title** | | **Source:** | **Convenor (Jens-Rainer Ohm)** | **Convenor (Jens-Rainer Ohm)** | | **Type:** | **General** | **General** | | **Subtype:** | **Other** | **N/A** | | **Status:** | **Approved** | **Approved** | | **Date:** | **2023-05-15** | **2023-04-28** | | **Expected Action:** | **Info** | **Info** | | **Action due date:** | **N/A** | **N/A** | | **No. of pages** | **12** (without this cover page) | **X** (without this cover page) | | **Email of convenor:** | **ohm @ ient . rwth-aachen . de** | **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** | **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**  30th Meeting, Antalya, TR, 21–28 April 2023 | Document: JVET-AD2023-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Exploration Experiments 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, D. Rusanovskyy, M. Santamaria, J. [Ström](mailto:jacob.strom@ericsson.com), 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)  dmytror@qti.qualcomm.com  [maria.santamaria\_gomez@nokia.com](mailto:maria.santamaria_gomez@nokia.com)  [jacob.strom@ericsson.com](mailto:jacob.strom@ericsson.com)  [liqiangwang@tencent.com](mailto:liqiangwang@tencent.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-AD and JVET-AE meetings to evaluate **Neural Network-based Video Coding (**NNVC) technologies, analyze their performance, and analyze their complexity aspects.

# Introduction

In JVET-AD meeting low complexity operation point NN-based filter was agreed to be default configuration of NNVC-5.0. Group keeps working on high operation point unified filter (UF) NN-based filter.

This round of EE1 will include EE1-0 test, which is an implementation and training from the scratch the unified filter design ([JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)). EE1-0 will be performed jointly and become comparison point for all EE1 tests targeting UF improvement.

EE1-1 category tests which investigate Unified Filter (UF) architecture change and require training shall be described using NN architecture description in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) (excel). It is strongly recommended to modify no more than one aspect in each EE1-1 sub-test.

EE1-1 category tests which investigate Unified Filter (UF) architecture recommended to start training in parallel with EE1-0 and requested to follow unified training strategy (described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944), copied here in section 11).

EE1-2 category tests which modify training procedure recommended to experiment with Unified Filter (UF) as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) and clearly describe changes compared to unified training procedure.

EE1-3 category tests which propose filter usage aspects (Table 2 in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)) or study content adaptation recommended to conduct test using Unified Filter (UF) as it will be developed in EE1-0 (w/o re-training models of EE1-0).

EE1-4 category tests are for modification of Low complexity filter in NNVC-5.0.

Tests on NN-based inter and NN-based intra technologies are EE1-5 and EE1-6 categories respectively.

Improvements for NN-based super resolution is category EE1-7.

All tests in EE1 advised to use NNVC5.0 as code base (unless it is not possible). The anchor for EE1 test is default configuration of NNVC5.0 as defined by AhG11/AhG14 (NN-intra and low complexity NN-filter expected to be enabled by default) in JVET-AD2016. Anchor performance to be provided by AhG14. Reference point for high operation point NN-filters is EE1-0.

Tests planned in this EE1 are summarized in Table 1.

Table 1 Overview of EE1 tests

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test** | **Architecture** | **Training** | **Usage in a codec** | **Tester/Trainer** | **Cross-check** |
| **EE1-0** | Unified filter (UF) as in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | | | Bytedance, Nokia, Qualcomm Tencent, OPPO | |
| EE1-1.1 | Separate Y and UV | as in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | | Bytedance | training +test |
| EE1-1.2 | Decomposed CONV | Qualcomm | training +test |
| EE1-1.3 | Filter “head” | Qualcomm | training +test |
| EE1-1.4 | 4 sub-tests: temporal filter, residual, filter “head”, back bone blocks design | Bytedance | training +test |
| EE1-1.5 | 3 sub-tests: filter “head”, residual block and split network design | Tencent | training +test |
| EE1-2 | [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | Training set | [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | Tencent | training +test |
| EE1-3.1 | [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | | flipping | OPPO | test only |
| EE1-3.2 | adjustment | OPPO | test only |
| EE1-3.3 | RDO | Bytedance | training +test |
| EE1-4.1 | [Improved](https://jvet-experts.org/doc_end_user/current_document.php?id=12708) Low OP | [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708) | as in NNVC-5.0 | Dolby/Ittiam | test only |
| EE1-4.2 | Simplified UF | [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | Bytedance | training +test |
| EE1-4.3 | Simplified UF | Tencent | training +test |
| EE1-4.4 | Simplified UF | Qualcomm | training+test |
| EE1-5.1 | NN-inter | | | Tencent/WHU | test only |
| EE1-6.1 | NN-intra | | | Inter Digital | training +test |
| EE1-7.1 | [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | JVET-AC0196, [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) | Super-resolution | Tencent | training +test |

# Unified Filter Training

## EE1-0 Agreed unified filter design for High Operation Point

Filter architecture: Unified filter (UF) architecture is described attached excel (page EE1-0) and depicted on Figure 1. Design follows decision of BoG ([JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)). The number of Back Bone blocks was selected to be 24 targeting complexity 477 kMAC/pxl.

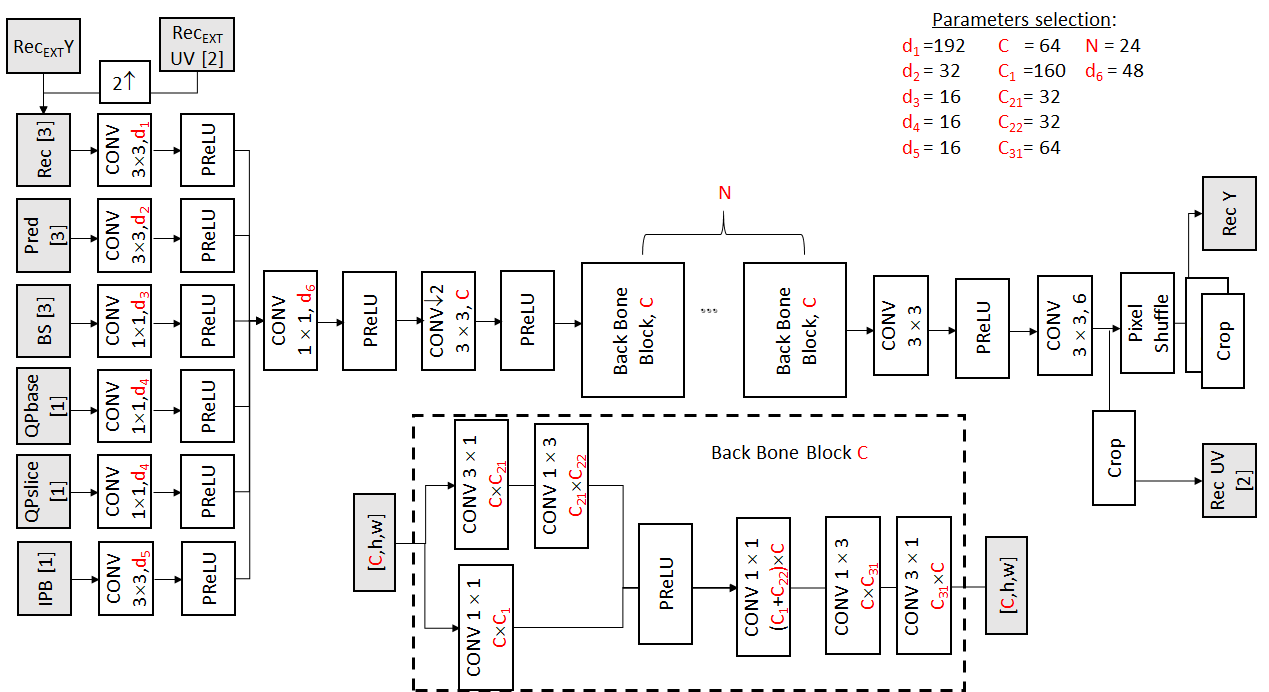


Figure Unified filter design for High Operation point

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Qualcomm, Bytedance, Tencent, Nokia, OPPO

Cross-check of training: multiple runs of training by different companies and testing cross-check after each training stage.

Notes: 1) Difference between different runs of training to be reported to the group and be considered as negligible performance deviation (gain below this deviation cannot be considered as argument for filter architecture modification). 2) EE1-0 is common reference point for all tests targeting improvement of high operation point NN-filter.

# Unified Filter Architectural changes

## EE1-1.1 Separate model for Luma and Chroma

Filter architecture: Unified filter (UF) architecture is described in the attached excel (page EE1-0) following decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944). This tests investigate separate architectures for luma and chroma models. To facilitate an easier comparison with UF, numbers of backbone blocks and channels for luma and chroma models are expected to be optimized, subject to giving a similar complexity as the UF (e.g. about 477 kMAC/pxl).

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Bytedance

Cross-check of training: Training cross-check is planned, cross-checker to be identified later.

Notes: Difference from EE1-0 (UF) is illustrated by Figure 2. This test targets to optimize separate model designs for luma and chroma, along methods of JVET-AD0106.

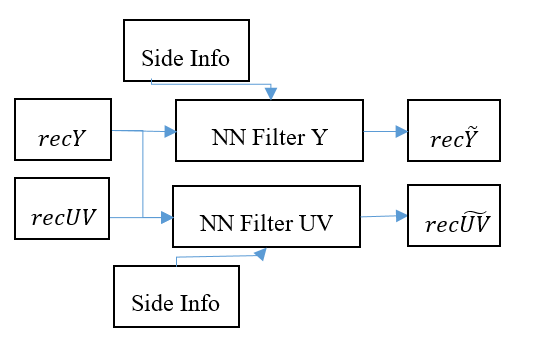
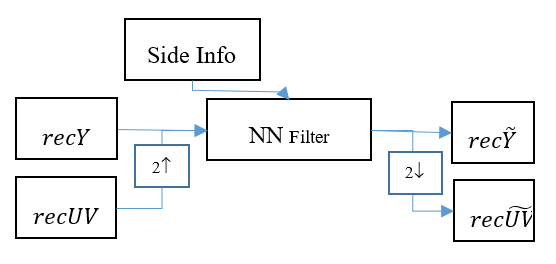


Figure Joint filtration of three color components in UF (left), separate filtration for Luminance and Chrominance (right).

## EE1-1.2 Complexity-performance tradeoff of decomposition

Filter architecture: NNVC-5.0, unified filter architecture.

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Qualcomm

Cross-check of training: Training cross-check is planned, cross-checker to be identified later

Notes: This test targets to optimize the complexity-performance trade-off of the convolution decomposition and residual block structures, along methods of EE1-1.3.5 (JVET-AD0205) and JVET-AD0211. The test will include sub-tests with across architecture complexity-performance optimization.

## EE1-1.3 Study on input feature set optimization

Filter architecture: NNVC-5.0, unified filter architecture.

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Qualcomm

Cross-check of training: TBA

Notes: This test targets to optimize input feature set extraction, along methods of EE1-1.3.2 (JVET-AD0205).

## EE1-1.4 Study on network architecture optimization

Filter architecture: NNVC-5.0, based on unified filter architecture.

Training strategy: based on unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Bytedance

Cross-check of training: Training cross-check is planned, cross-checker to be identified later

Notes: This test will include subtests targeting to optimize the different aspects of network architecture design, e.g. inputs, transformer block, residual block, layer decomposition, split architecture, temporal filter, etc., along methods of JVET-AC0177, JVET-AD0106, JVET-AD0237.

## EE1-1.5 Optimization for complexity-performance trade-off of network

Filter architecture: NNVC-5.0, unified filter architecture.

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

Training to be conducted: Tencent

Cross-check of training: Training cross-check is planned, cross-checker to be identified later

Notes: This test will include subtests targeting to optimize the complexity-performance trade-off of network architecture, such as inputs, residual block structure and split network architecture, along methods of JVET-AD0167.

## EE1-2.1 Simplification for training dataset

Filter architecture: Unified filter (UF) architecture is described attached excel (page EE1-0) follow decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944).

Training strategy: Unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944) except training dataset aspect

Training to be conducted: Tencent

Cross-check of training: TBA.

Notes: This test target to simplify the training dataset while maintaining complexity-performance trade-off.

# Unified filter usage aspects

## EE1-3.1 Flipping of input and output of model in UF(JVET-AD0069)

Filter architecture: Unified filter (UF) architecture is described attached excel (page EE1-0) follow decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944).

Training strategy: no re-training

Training to be conducted: no training

Tester: OPPO

Cross-check of test: Cross-check is planned, cross-checker to be identified later.

Notes: Flipping operations applied to the inputs and output of the UF neural network are proposed to further improve the performance in this contribution. The input of UF includes reconstructed samples, prediction samples, BS, BaseQP, SliceQP and IPB. When a flipping operation is applied, only the reconstructed samples, prediction samples, BS, and IPB are flipped and fed into the network with the other inputs to perform the inference. Then the output of network is flipped back to restore the order of the filtered samples. Both horizontal and vertical flip can be applied to input and output of network.

## EE1-3.2 On adjustment of residual for UF(JVET-AD0068)

Filter architecture: Unified filter (UF) architecture is described attached excel (page EE1-0) follow decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944).

Training strategy: no re-training

Training to be conducted: no training

Cross-check of test: Cross-check is planned, cross-checker to be identified later

Notes: This test evaluates the residual offset adjustment and combination with chroma order adjustment on top of UF. A frame level residual offset can be used to adjust the output of the NNLF in the inference stage. More specifically, the residual is adjusted by reducing the magnitude of the residual at each pixel by a small offset value, and the offset candidates are {1, 2}. Also, a frame level chroma order adjustment method can be used to allow the input/output order switch between the U and V components of the neural network-based loop filters in the inference stage.

## EE1-3.3 RDO filter of UF

Filter architecture: Unified filter (UF) architecture is described in the attached excel (page EE1-0) following decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944). RDO filter architecture will be designed based on the UF. Specifically, number of layers and feature maps will be reduced for the RDO model. The optimal numbers are expected to be figured out in this EE test.

Training strategy: based on unified training as described in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944).

Training to be conducted: Bytedance

Cross-check of training: Cross-check for training and test is planned, cross-checker to be identified later.

Notes: This tests target to optimize the design of the RDO model, e.g. layer numbers, channel numbers, etc., along methods of JVET-AD0189.

# Improvement for Low complexity NN-filter

## EE1-4.1 Neural-network loop filters with further complexity reduction

Filter architecture: Filter architecture is described attached excel (page EE1-4.1) follow decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944). The model uses CP decomposition, fusing adjacent 1x1 convolution and split architecture for luma and chroma, as shown in Figure 3, which is explained in detail in [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708).

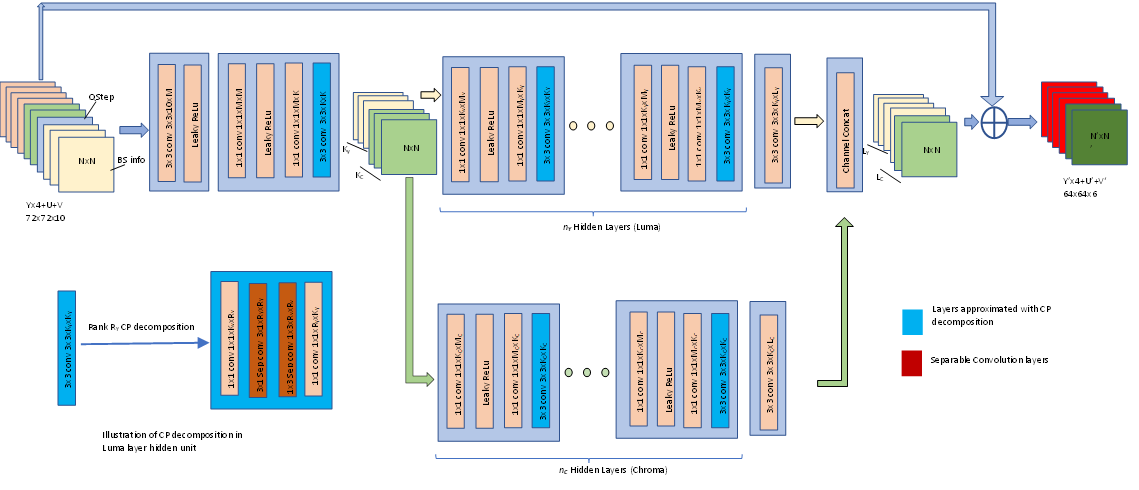


Figure Split luma and chroma model + CP Decomposition + fusion of 1x1 conv layer

Training strategy: use the training strategy described in [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708).

Training to be conducted: perform trainings with different settings to explore the possibility of further complexity reduction with the following aspects:

* Different combinations of model parameters
  + Number of hidden layers
  + Number of channels in luma and chroma paths
* SIMD friendly sizes: number of channels preferably multiple of 16 (8 also possible but slower)

Cross-check of inference: Ericsson

## EE1-4.2 Neural-network loop filters with further design optimization

Filter architecture: Filter architecture will be designed based on the one described in the attached excel (page EE1 (LOP)) following decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944). It is proposed to optimize the filter design by borrowing elements from its counterpart in high operation point.

Training strategy: based on the training strategy described in [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708).

Training to be conducted: Bytedance

Cross-check of training: Training cross-check is planned, cross-checker to be identified later.

Notes: perform trainings/inferences with different settings to explore the possibility of further design optimization.

## EE1-4.3 Optimization for neural-network loop filters based on high operation point and EE1-1.5

Filter architecture: Filter architecture is described attached excel (page EE1 (LOP)) follow decision in [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944). It is proposed to optimize the trade-off by using elements from high operation point.

Training strategy: training strategy described in [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708), JVET-AD0380.

Training to be conducted: Tencent

Cross-check of training: Training cross-check is planned, cross-checker to be identified later.

Notes: Further training/inference optimization for low-complexity neural-network loop filters by considering the effective elements in high operation point and related EE1 tests.

## EE1-4.4 Low complexity NN filter with design elements of UF and EE1-1.2 and EE1-1.3

Filter architecture: Filter architecture of EE1 (LOP) with design elements of Unified Filter architecture, EE1-1.2 and EE1-1.3.

Training strategy: training strategy described in [JVET-AD0157](https://jvet-experts.org/doc_end_user/current_document.php?id=12708), JVET-AD0380.

Training to be conducted: Qualcomm

Cross-check of training: Training cross-check is planned, cross-checker to be identified later.

Notes: Optimization of low complexity NN filter architecture and training process with elements of UF and related EE1 tests.

# NN-based Inter

## EE1-5.1 NN-based inter prediction with lower complexity

Filter architecture: Figure 4illustrates the network architecture of the inter frame generation network, which is based on the model described in JVET-AD0162. To mitigate complexity, the number of ResBlocks and convolution channels will be reduced.

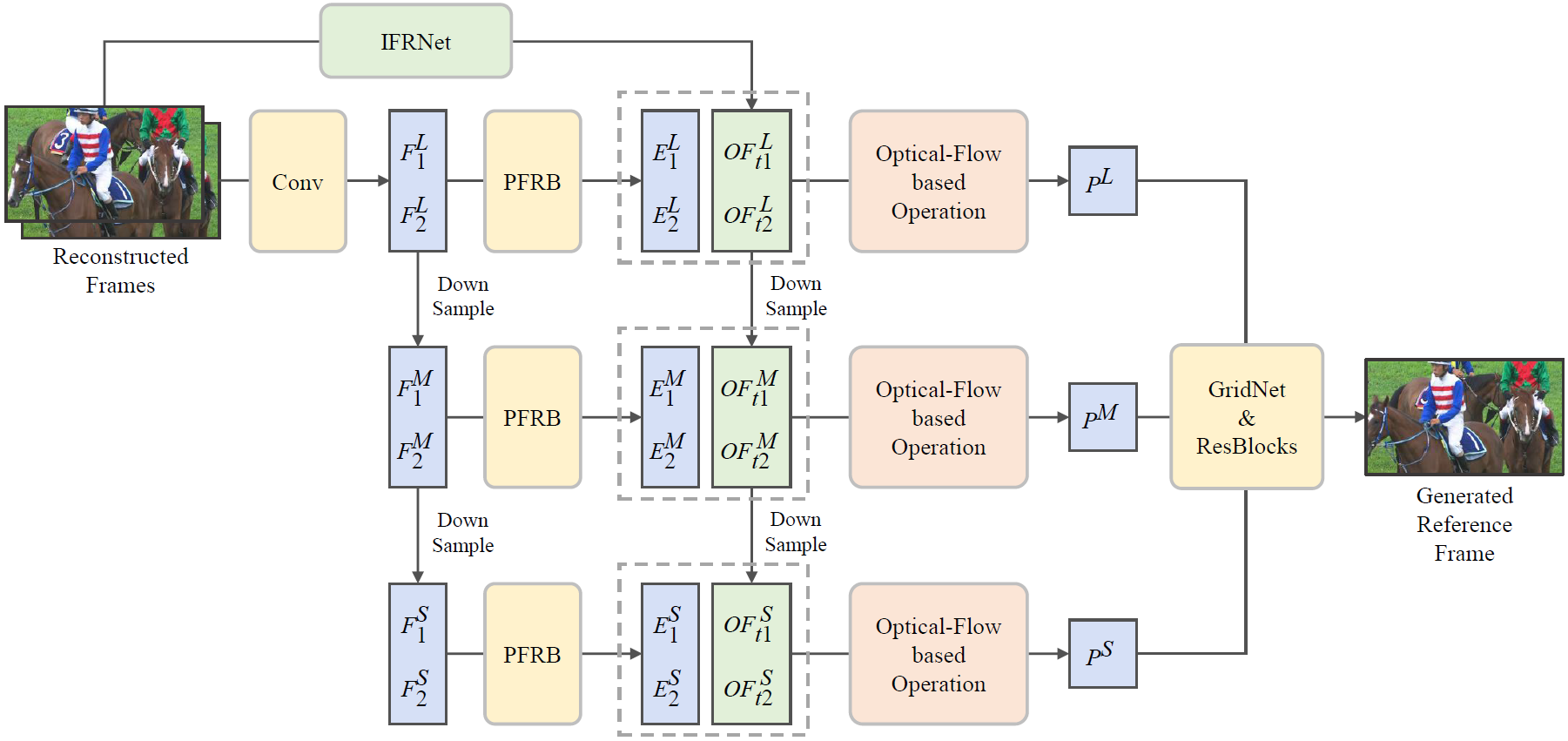


Figure The network architecture of the inter frame generation method

Training strategy: Use the compressed Vimeo90K dataset to train the model. The pictures in the dataset are compressed by VTM, with QP value randomly selected from 22, 27, 32, 37, and 42. The loss function is L1 loss, which is minimized using the Adam optimizer. A pre-trained IFRNet is used as our flow estimator, with a learning rate set at . The other parts of the network trained at a learning rate of . The training epochs are set to 200.

Training to be conducted: Wuhan University, Tencent.

Cross-check of training: Not planned. Inference cross-check only.

Notes: This test aims to reduce the complexity of the NN-based inter method based on JVET-AD0162 (EE1-2.1-related: DRF Model without QP Input).

# NN-based intra

## EE1-6.1 Neural network-based intra prediction with reduced complexity

Filter architecture: the neural network-based intra prediction mode (Figure 5) gathers 7 neural networks, each predicting blocks of a different size in . The architecture of the neural network predicting blocks of size is shown in the figure below, see also JVET-AD0212.

Diagram

Description automatically generated

Figure 5: architecture of the neural network predicting blocks of size , belonging to the neural network-based intra prediction mode.

|  |  |
| --- | --- |
| **architecture specificity** | **JVET-AD0212** |
| Number of hidden layers | For 16x16, 3.  For the other 6 neural networks, 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 4 neural networks, 1216. |
| Biases | Only for the layer returning and the layer returning and . |
| Non-linearity | LeakyReLU in all hidden layers. |

Table 2: architecture specificities for each neural network belonging to the neural network-based intra prediction mode.

Training strategy: iterative training comprising 4 cycles, as described in JVET-AC2019.

Training to be conducted: iterative training from scratch of the neural network-based intra prediction mode, involving the architecture and weight sparsity changes detailed in JVET-AD0212.

Cross-check of training: to be added soon.

# NN-based super resolution

## EE1-7.1 Neural network based super resolution based on the unified filter architecture

Filter architecture: NNVC-5.0, unified filter architecture.

Training strategy: training strategy described in JVET-AC0196, JVET-AD0380.

Training to be conducted: Tencent

Cross-check of training: TBA.

Notes: This test target to optimize the super resolution network based on the unified architecture and training strategy.

* Train the super-resolution network based on the unified filter solution
* Train the super-resolution network with enabling NN-intra and NN-filter

# Timeline

**T1 - 2 weeks after JVET-AD meeting (12-May-2023):** To revise EE description. Changes should be discussed and agreed on JVET reflector.

**T2 - 3 weeks after JVET-AD meeting (19-May-2023):** (in coordination with AhG14) SW for UF (EE1-0) training is ready, announced and joint training starts. Anchors (including technologies in NNVC-5.0) are available.

**T3 – 5 weeks after JVET-AD meeting (02-June-2023)**: Software release for tests with training cross-check. Sufficient explanation about training scripts has been provided to the cross-checker. Training verification starts.

**T4 – 7 weeks after JVET-AD meeting (16-June-2023)**: (in coordination with AhG14) AhG11/AhG14 teleconference for UF (EE1-0) training status discussion.

**T5 – 1.5 week before T7 (30-June-2023):** Software release for tests w/o training cross-check. Test verification starts.

**T6 – 3 days before T7 (08-July-2023):** Cross-checkers report status to EE1 coordinators (sending e-mail).

**T7 – 11-July-2023:** EE1 summary is uploaded as input contribution

# SW location

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

# References

## [1] [E. Alshina](mailto:elena.alshina@huawei.com), [F. Galpin](mailto:franck.galpin@interdigital.com) BoG report on NN-filter design unification [JVET-AD0380](https://jvet-experts.org/doc_end_user/current_document.php?id=12944)

[2] E. Alshina, R.-L. Liao, S. Liu, A. Segall Common test conditions and evaluation procedures for neural network-based video coding technology [JVET-AD2016](https://jvet-experts.org/doc_end_user/current_document.php?id=12978)

# Unified filter training

Training description follow agreement reached in [1]

1. Training Model stage I
   1. Data extraction for intra from vanilla VTM
      1. encoder turns on md5sum, just NNVC-4.0 all NNVC tools disabled, encoder\_intra\_vtm.cfg, just first frame
      2. QPs = 22,27,32,37,42 + 19,24,29,34,39
      3. DIV2K (4:4:4 RGB 🡪 YUV420 10 bits, version ffmpeg to be specified, script is part of NNVC SW)
      4. Rotation, flipping, down-sampling (provided by ByteDance + list of output + md5sum)
      5. Reconstruction is extracted before de-block
      6. All CTUs are extracted (except those on the border) 🡪 list of md5sums (NNVC SW coordinators to provide) 🡪 10 days
      7. Status check via e-mail
   2. Model stage I training from scratch for Intra data:
      1. Random initialization (NNVC SW decide)
      2. Use all CTUs
      3. Learning Rate scheduler to be provided by Qualcomm
      4. Resume of training (track of learning rate, store check-points…, must be added to the training scripts) functionality is in training scripts to be adjusted for unified model
      5. Status check via e-mail (learning curve) 🡪 10 days
      6. *🡪 RA and all intra cfg results under AhG11 test condition (model is used only for Intra slice)*
2. Training Model stage II
   1. Data extraction with model#1:
      1. encoder turns on md5sum, NNVC-4.0 + Model stage I for I-frames only, encoder\_ra\_model1.cfg, 65 frames (last is duplicated), Intra Period 64, GOP=32.
      2. QPs = 22,27,32,37,42
      3. List of sequences
         1. TVD: 65 frames, 20 sequences (Tencent provides list of sequences with md5sum)
         2. BVI (YUV420), 620 sequences (Bytedance provides list of sequences with md5sum + script for duplication last frame duplicated)
         3. All reconstructed CTU are extracted before de-block , list of frames to be provided by Tencent and Bytedance (after check) + DVI2K training data as in training stage I
      4. Status check via e-mail 🡪 10 days
   2. Model stage II training with extracted data:
      1. Random initialization (NNVC SW decide)
      2. Use all CTUs
      3. Rotation and flipping is part of training already
      4. Learning Rate scheduler to be provided by Qualcomm
      5. Resume of training (track of learning rate, store check-points…, must be added to the training scripts) functionality is in training scripts to be adjusted for unified model
      6. *🡪 RA and all intra cfg results under AhG11 test condition*
      7. Status check via e-mail (learning curve) (likely telco will be needed) 🡪decide next stage 🡪 10 days
3. Training stage III
   1. Data extraction with Model stage II:
      1. encoder turns on md5sum, NNVC-4.0 + Model stage II (both Intra and Inter frames), encoder\_ra\_model1.cfg, 65 frames (last is duplicated), Intra Period 64, GOP=32.
      2. QPs = 22,27,32,37,42
      3. List of sequences
         1. TVD: 65 frames, 20 sequences (Tencent provides list of sequences with md5sum)
         2. BVI (YUV420), 620 sequences (Bytedance provides list of sequences with md5sum + script for duplication last frame duplicated)
         3. ALL reconstructed CTU are extracted before de-block , list of frames to be provided by Tencent and Bytedance (after check) + DVI2K training data as in training stage I
      4. Status check via e-mail 🡪 15 days
   2. Model *Stage III* training with extracted data:
      1. Random initialization (NNVC SW decide)
      2. Use all CTUs
      3. Rotation and flipping is part of training already (on the fly augmentation)
      4. Learning Rate scheduler to be provided by Qualcomm
      5. Resume of training (track of learning rate, store check-points…, must be added to the training scripts) functionality is in training scripts to be adjusted for unified model
      6. *🡪 RA and all intra cfg results under AhG11 test condition*
      7. Status check via e-mail (learning curve) + telco 🡪 15 days