

Overview of technologies considered in JVET's neural network-based video coding exploration

Andrew Segall

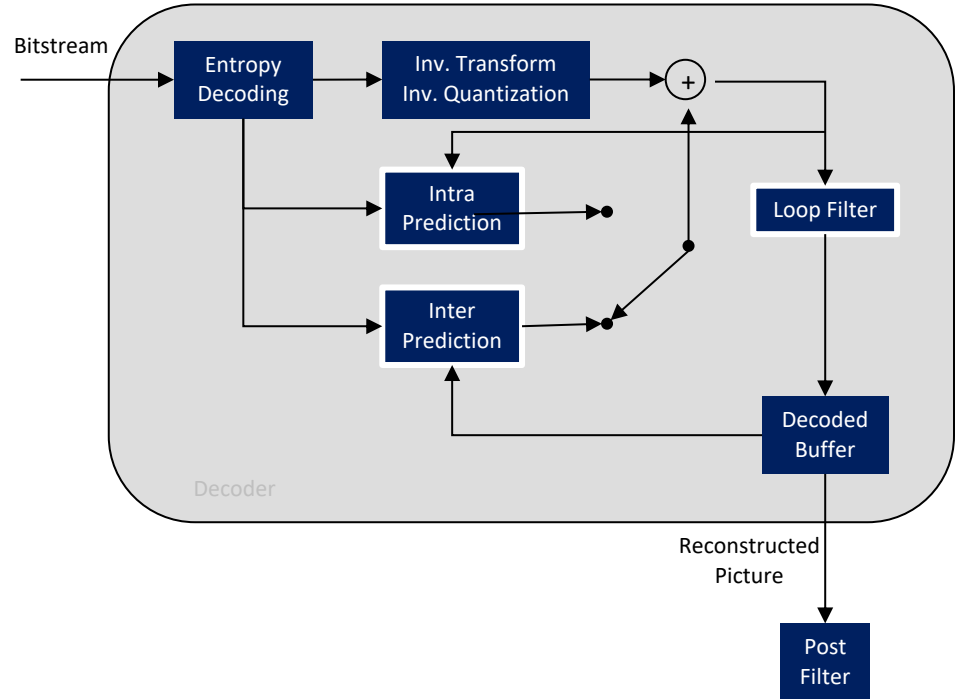


Introduction

- The Joint Video Experts Team (JVET) began formal exploration experiments studying neural network-based video coding in October 2020.
- Interest in this activity has been consistent, and JVET has received more than 80 technical inputs related to the topic in 2021.
- The goal of this presentation is to give an overview of technologies being proposed and studied.

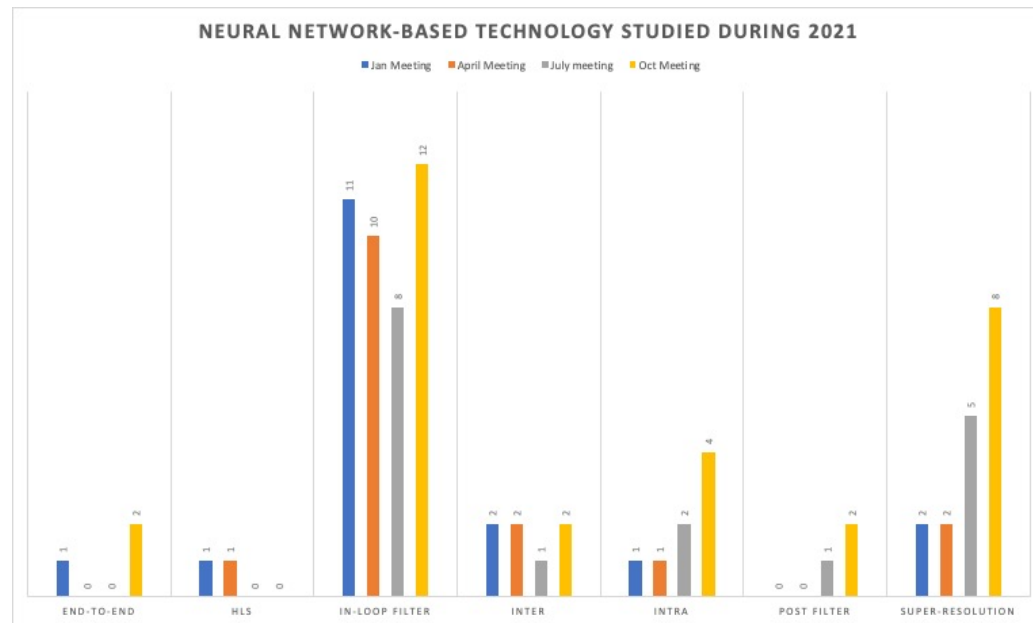
Categorization

- Organizing the survey is helped by categorizing the technologies
- Three major areas of work in JVET:
 1. Replacing existing VVC coding tools (or enhancing them)
 2. Introducing post-filtering operations
 3. End-to-end solutions that do not rely on VVC
- Tool replacement (or enhancement) is primarily focused on:
 1. In-loop filtering
 2. Super-resolution
 3. Intra-Prediction
 4. Inter-Prediction



Categorization

- Focus of JVET activity
 - Activity in all categories
 - Largest number of contributions are in:
 - In-loop filtering
 - 41 contributions
 - ~50% of technical input
 - Super-resolution
 - 17 contributions
 - Additional observation
 - Technical overlap between in-loop filter, post filter and super-resolution methods



Overview

- Overview
 - The rest of the presentation will be organized as follows:
 - Filtering methods
 - In-loop filtering
 - Post-Filtering
 - Super-Resolution
 - Intra-prediction methods
 - Inter-prediction methods
 - End-to-end methods
 - NN environments
 - Conclusions and more information

Filtering Methods: In-Loop Filtering

Loop Filtering

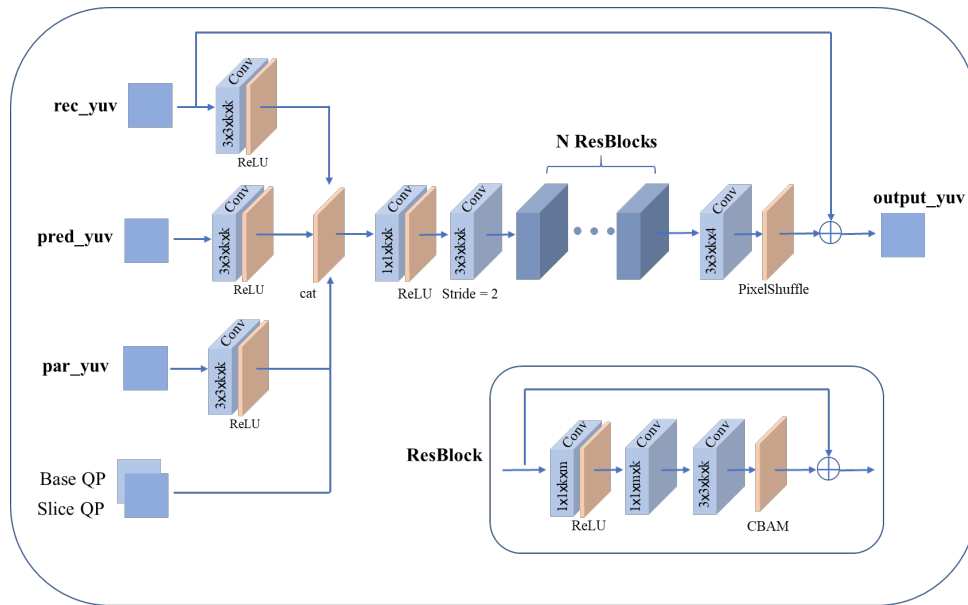
- Example of Loop Filtering Technology

- Inputs

- Reconstructed samples
 - Predicted samples
 - Partition parameters
 - Quantization parameters

- Common features

- Sequence of CNNs
 - Residual blocks
 - Shorter and longer skip connections
 - Attention mechanisms



JVET-X0055

$N = 8$

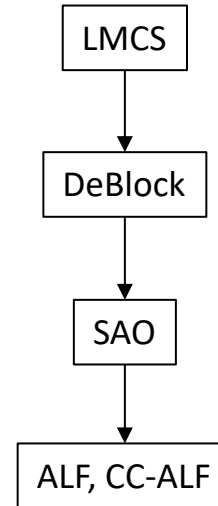
Channels = 64

Patch size = 144×144

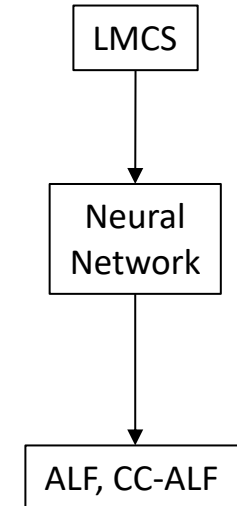
Loop Filtering

- Loop Filtering
 - Location of the neural network within the loop filter is an area of exploration
 - VVC design contains four major in-loop blocks
 - LMCS (Luma Mapping and Chroma Scaling)
 - Deblocking
 - SAO (Sample Adaptive Offset)
 - ALF (Adaptive Loop Filter)
 - Investigations
 - Adding neural networks to the in-loop process
 - Replacing elements of the in-loop blocks

VVC Loop Filter

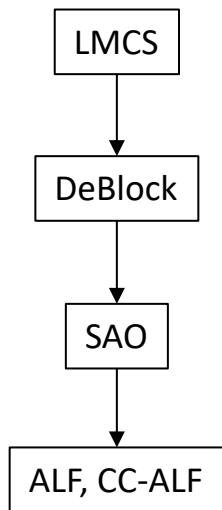


Example NN Loop Filter

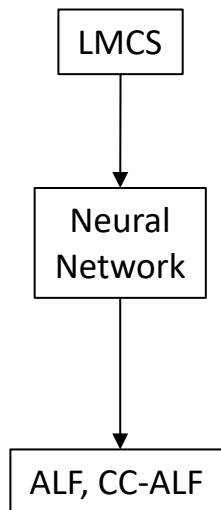


Loop Filtering

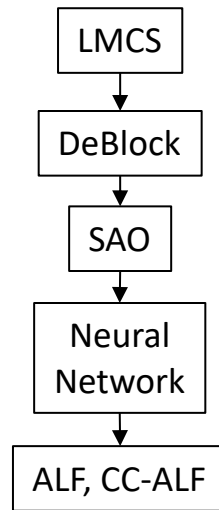
VVC Loop Filter



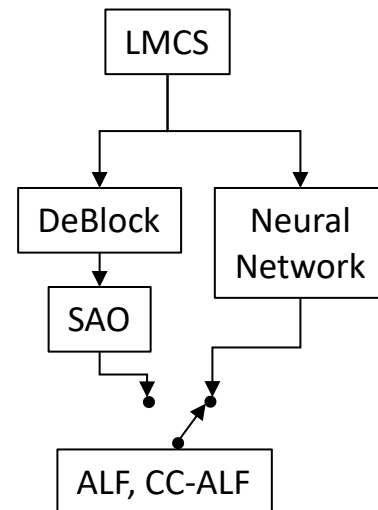
Example NN Loop Filter



Example NN Loop Filter

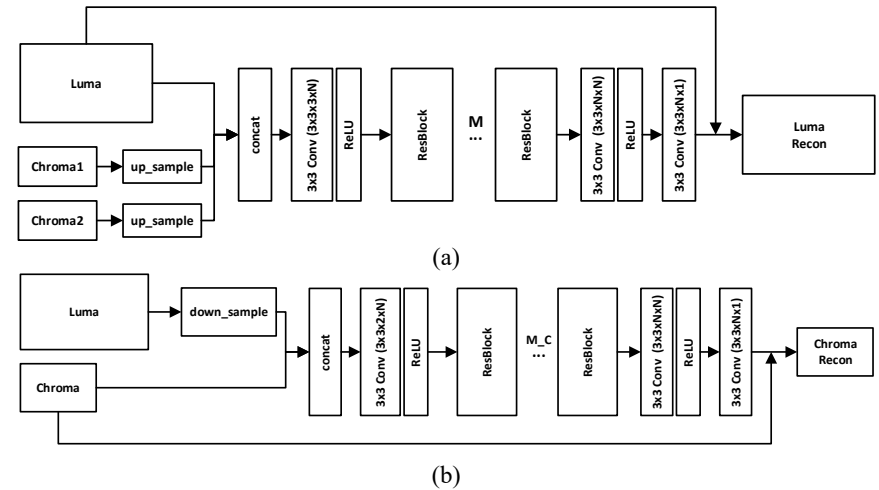


Example NN Loop Filter



Loop Filtering

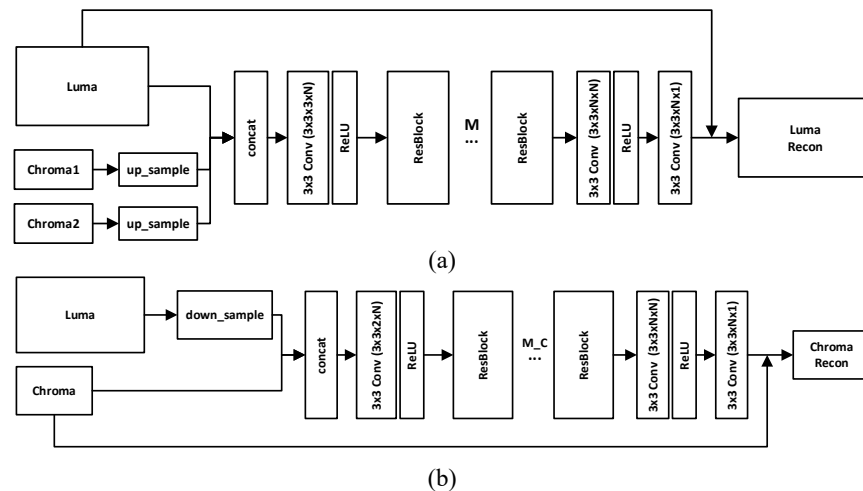
- Additional aspects under study
 - Benefit of cross channel correlation
 - Model parameter switching
 - Networks generally pre-trained with different weights loaded depending on conditions, such as:
 - Slice type
 - Chroma type
 - Base QP



JVET-X0094

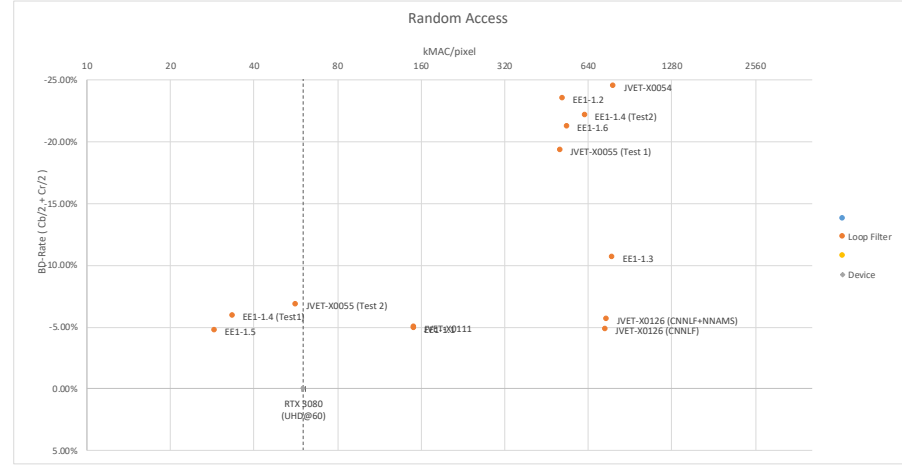
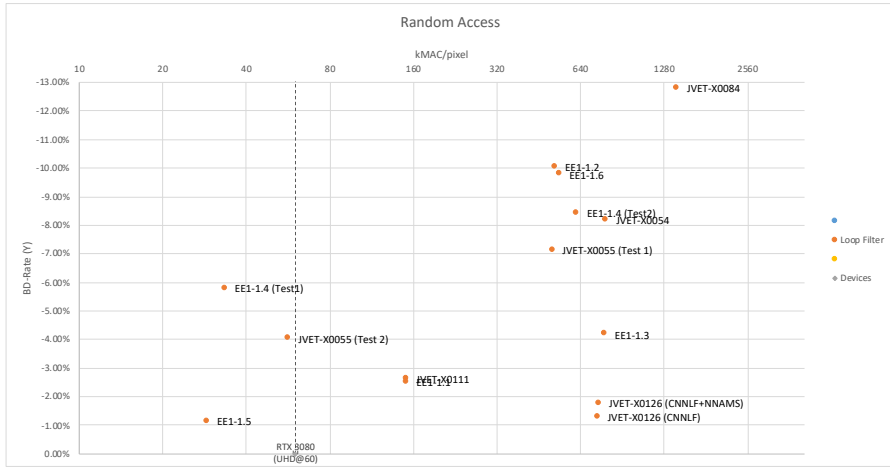
Loop Filtering

- Additional aspects under study
 - Switching granularity
 - Enable network per-frame
 - Enable network per-slice
 - Enable network per-block
 - Signal from candidate set
 - Switching precision
 - Binary switching
 - Signaling residual scaling factors

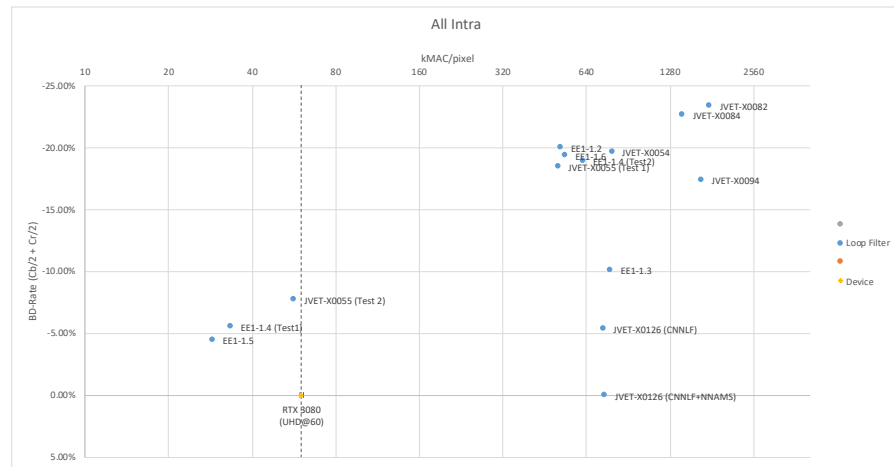
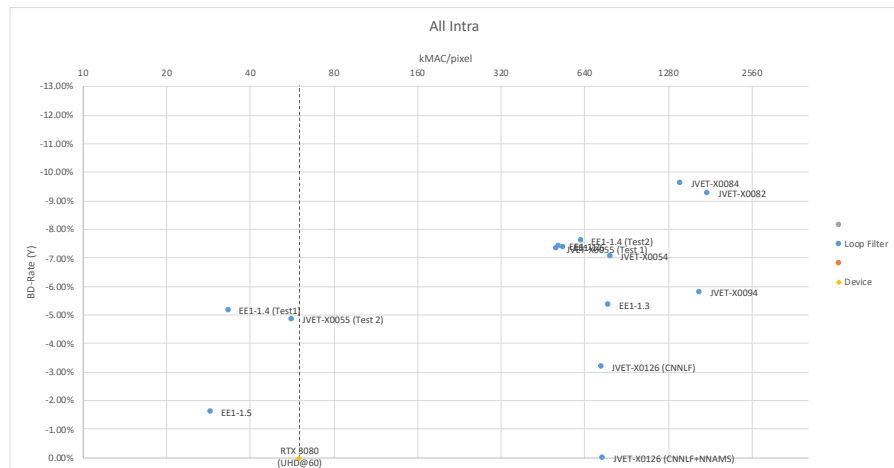


JVET-X0094

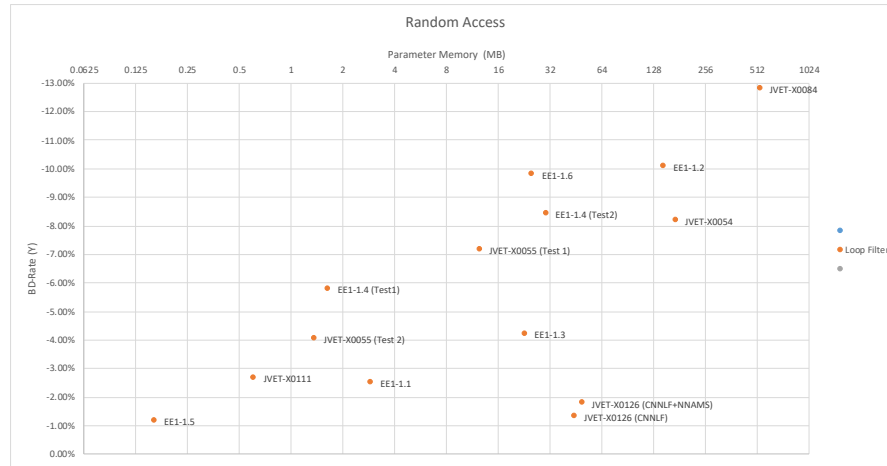
Loop Filtering



Loop Filtering



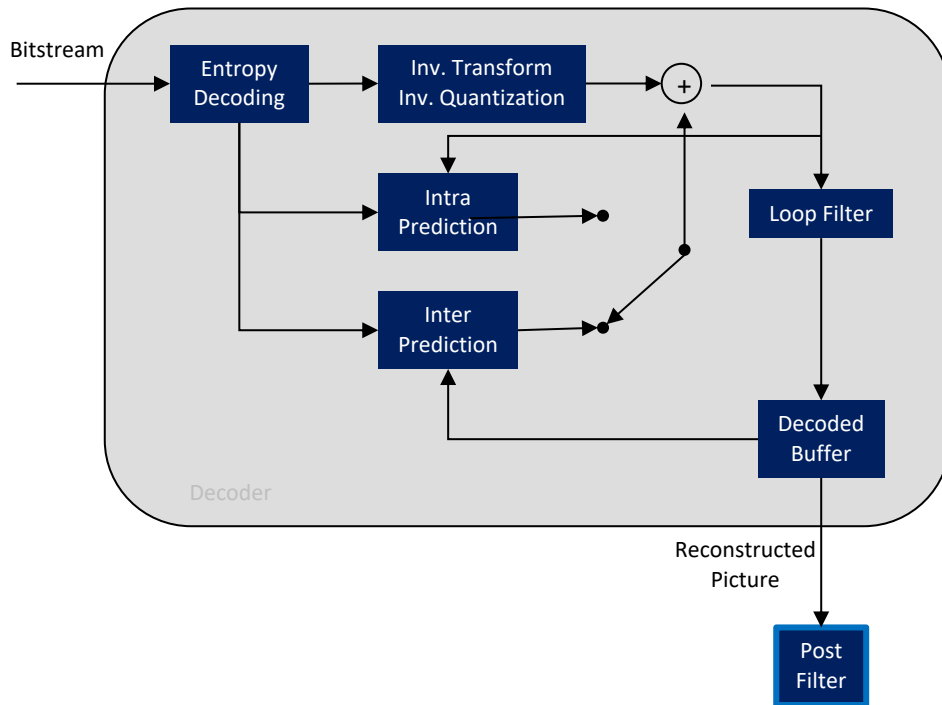
Loop Filtering



Filtering Methods: Post Filtering

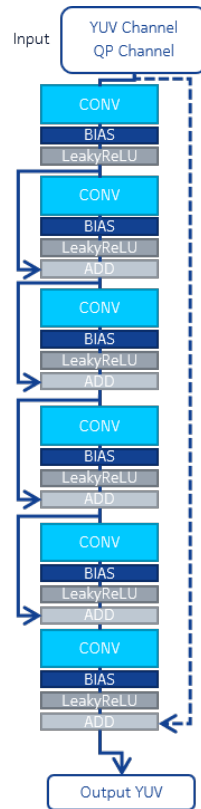
Post-Filtering

- Post-Filtering
 - Similar to loop filtering technology
 - Major conceptual difference is that the neural network is placed outside the coding loop



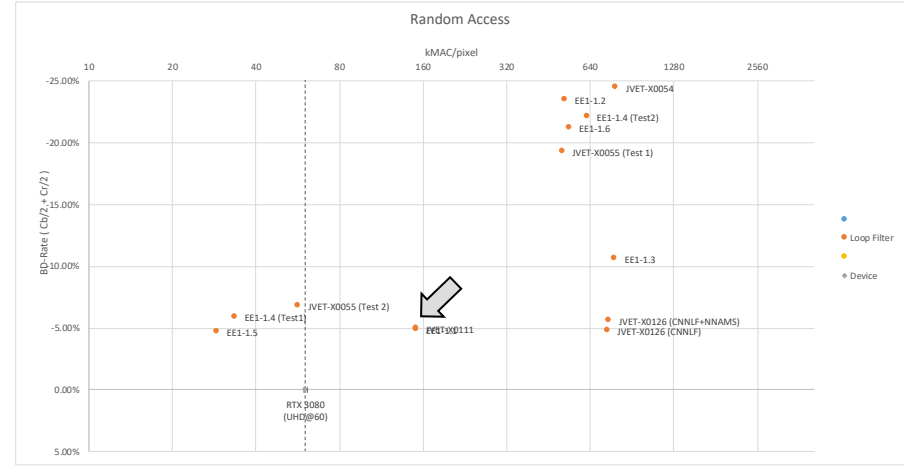
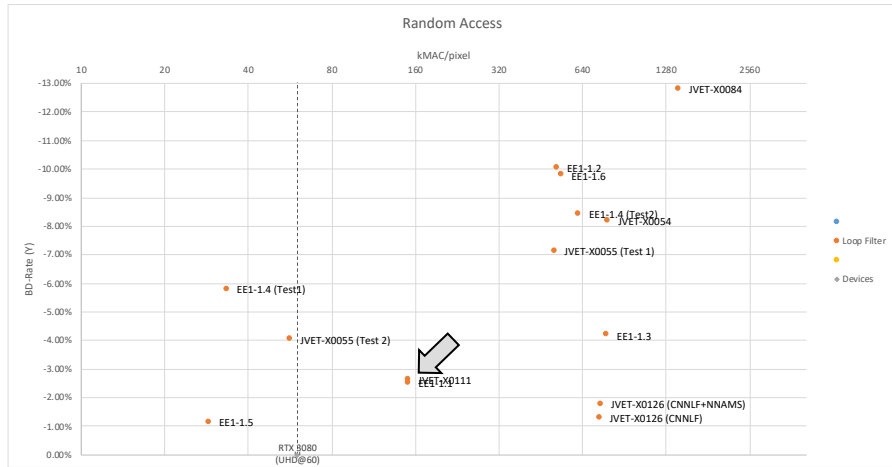
Post-Filtering

- Post-Filtering
 - Example technology
 - Sequence of CNNs
 - ResNet blocks
 - Shorter and longer skip connections
 - Reconstructed frame and QP information as input
 - Additional features
 - Model is pre-trained
 - Scaling parameters transmitted in bit-stream
 - Bias terms are refined and transmitted in the bit-stream

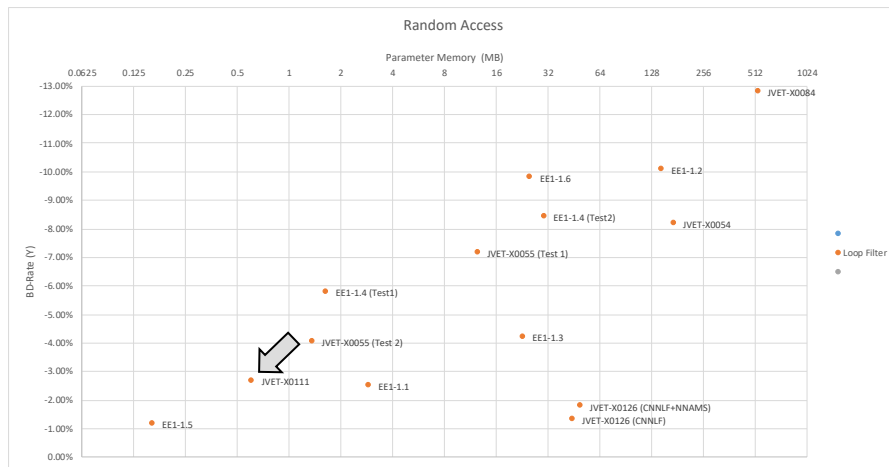


JVET-X0110

Post-Filtering



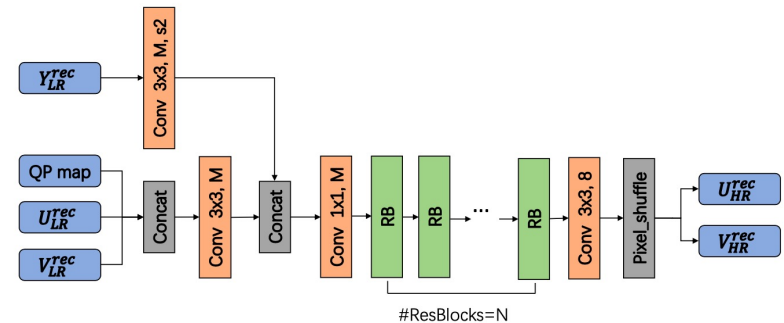
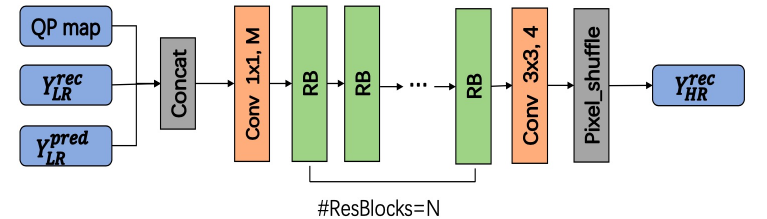
Post-Filtering



Filtering Methods: Super-Resolution

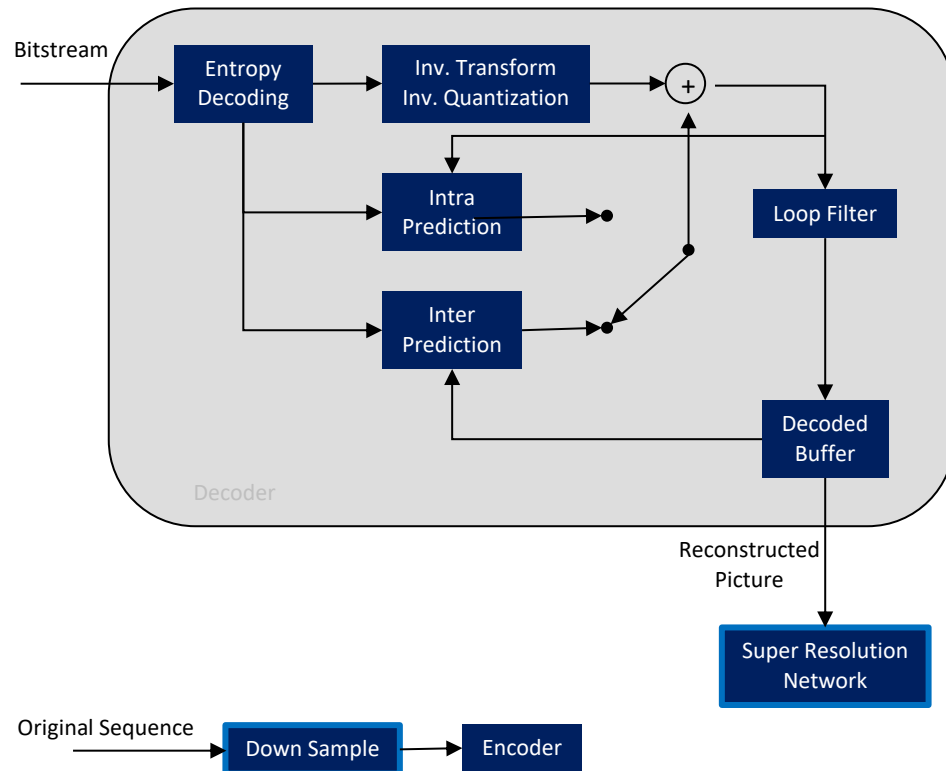
Super-Resolution

- Super-resolution
 - Viewed as a replacement of Reference Picture Resampling (RPR) in VVC
 - RPR allows for changing the resolution of the sequence within a bit-stream
 - Neural network solutions are proposed to replace the RPR operation
- Features are similar to in-loop filtering and post-processing approaches under study
 - Sequence of CNNs
 - ResNets
 - Longer term skip connections and attention less common (currently)



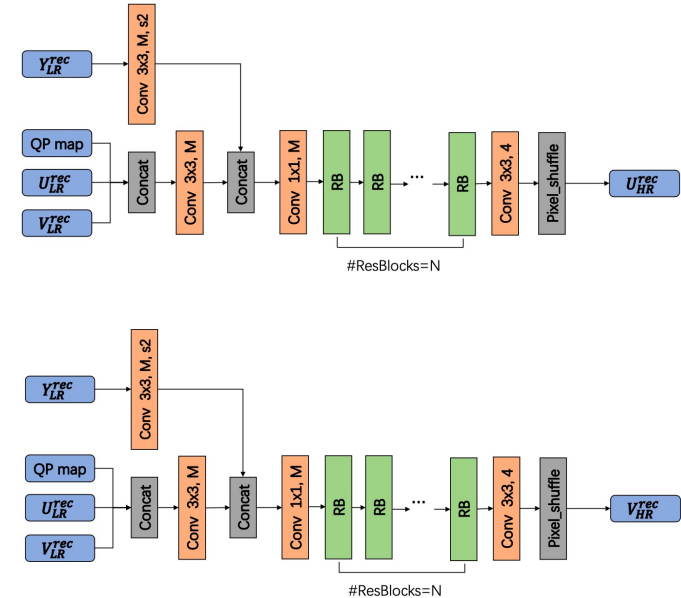
Super-Resolution

- Network placement
 - Super-resolution experiments do not incorporate switching resolution within a sequence
 - In practice, this means that the super-resolution network are located as a post-filter
 - Input sequence is down-sampled and compressed using VVC
 - Reconstructed sequence is upsampled using super-resolution network



Super-Resolution

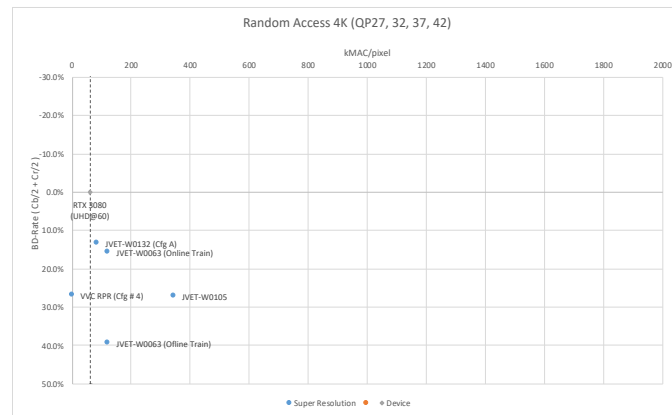
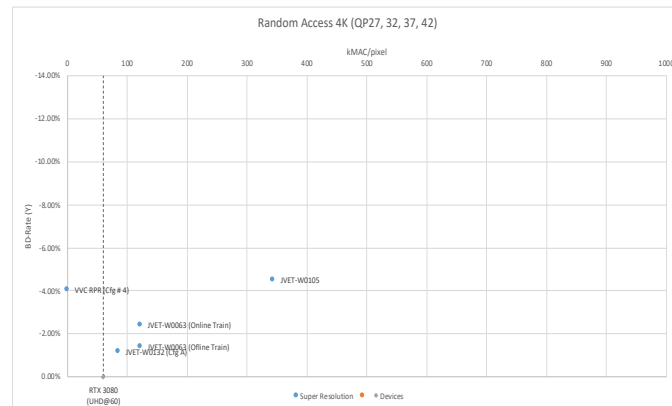
- Additional aspects under study
 - Use of different networks for chroma channels
 - Luma-chroma balance
 - Number of parameters and networks
 - Resolution control
 - Dynamic down-sampling to maximize coding efficiency
 - Fixed down-sampling for backwards compatibility constraint (e.g., 8K support)



Super-Resolution

■ Performance

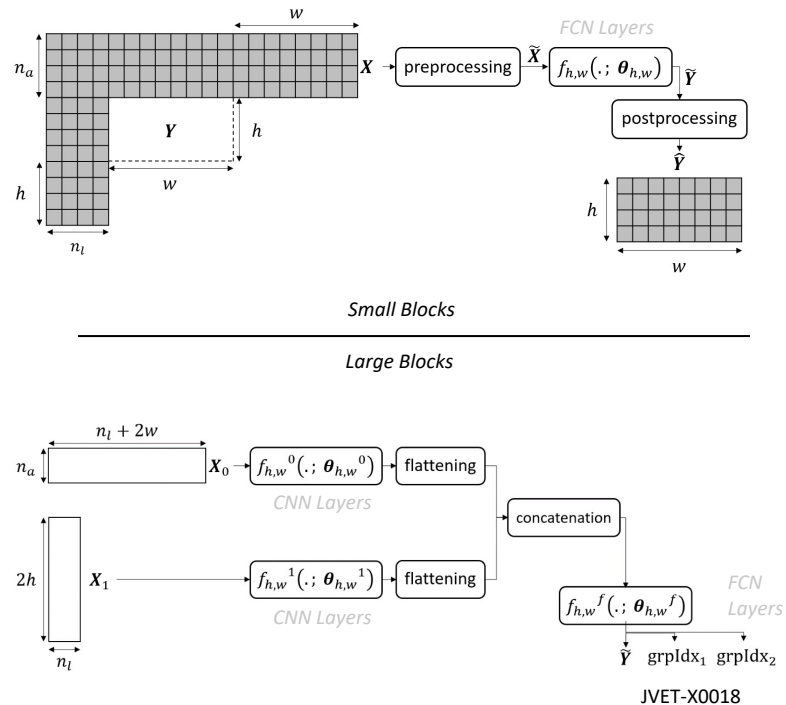
- Observation that super-resolution techniques primarily work at higher resolutions and/or lower bit-rates
- Majority of data provided to JVET focuses on low bit-rate 4K/UHD applications
 - July meeting
 - Luma gain of 4%
 - Chroma loss of 40%
 - October meeting
 - Average chroma loss decreased but still isolated issues



Intra-Prediction

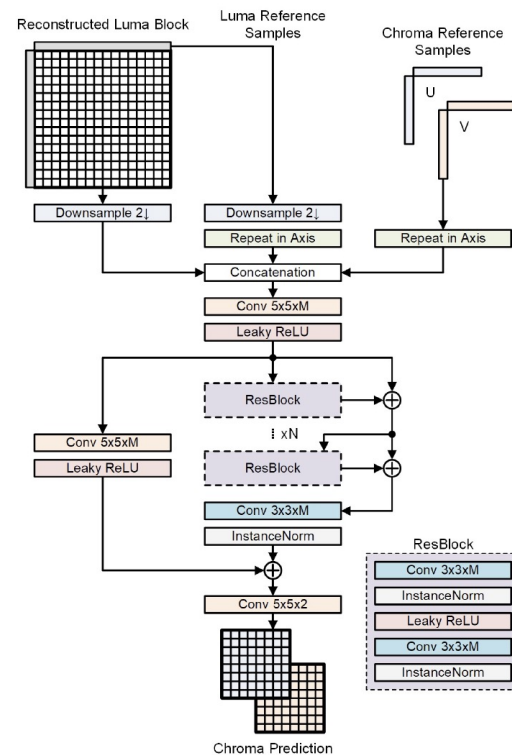
Intra-Prediction

- Intra-Prediction (Example)
 - Predict block Y
 - Small block sizes
 - $\min(h, w) \leq 8$
 - Three fully connected networks
 - $n_a = n_l = \min(h, w)$
 - Large blocks sizes
 - Convolutional network with six layers
 - $n_a = h/2; n_l = w/2$
 - Note: Prediction of grIdx (used for subsequent VVC processes)
- Recent emphasis has been on reducing complexity



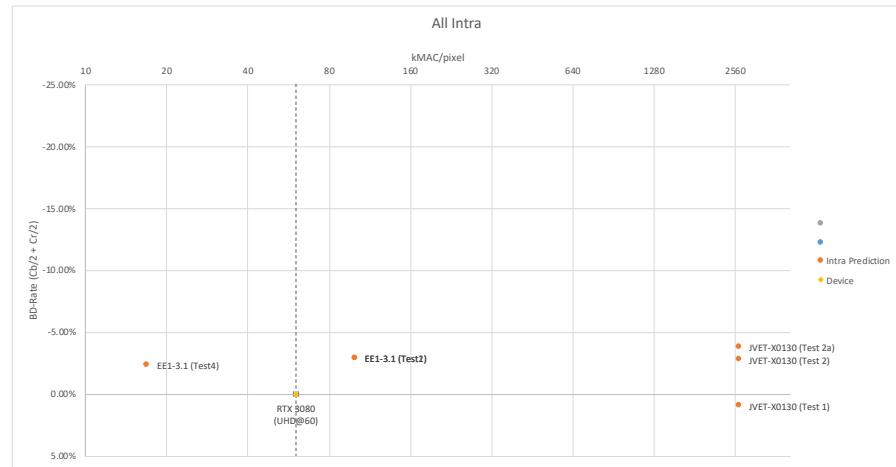
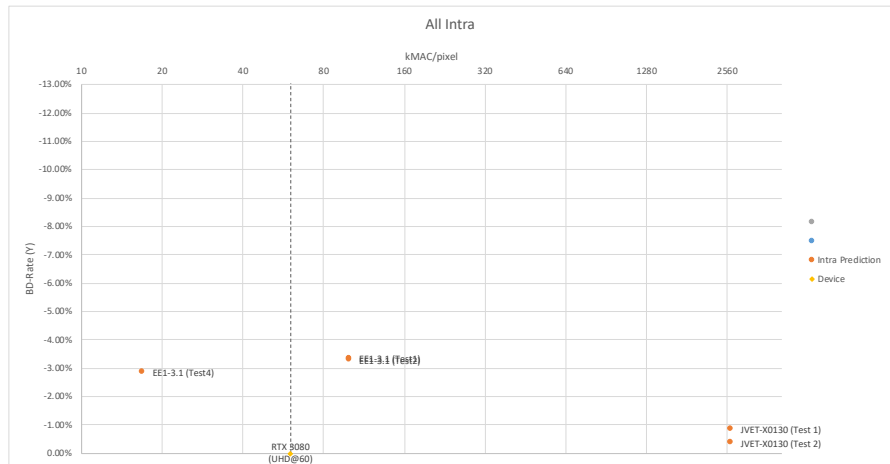
Intra-Prediction

- Additional Aspects
 - Cross-component linear model (CCLM) prediction
 - Goal: Improve/replace existing CCLM mode of VVC
 - Use of Auto-encoders for intra-prediction
 - Feature(s) extracted during encoding stage transmitted to decoder
 - Prediction and residual process replaced by auto-encoder
 - Auto-encoder added as a VVC mode (existing modes remain)



JVET-X0130

Intra-Prediction



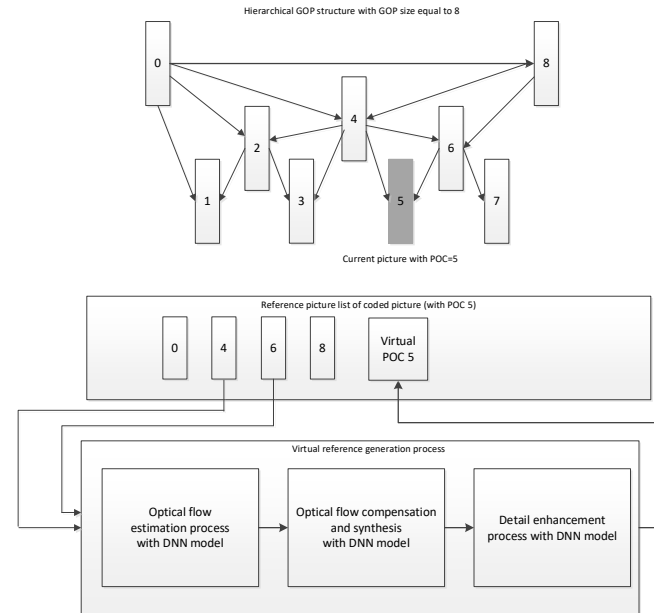
Intra-Prediction



Inter-Prediction

Inter-Prediction

- Virtual reference frames
 - Goal: Use neural network to synthesize a reference frame for prediction
- Current emphasis
 - Study benefit of detail enhancement block
 - Reported results
 - 2k MAC/sample
 - BD-Rate gains (Random Access)
 - -3.14% Class B (HD)
 - -4.69% Class C
 - -6.87% Class D



JVET-V0090

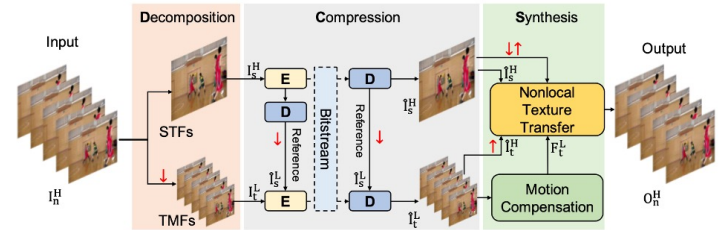
End-to-End

End-to-End

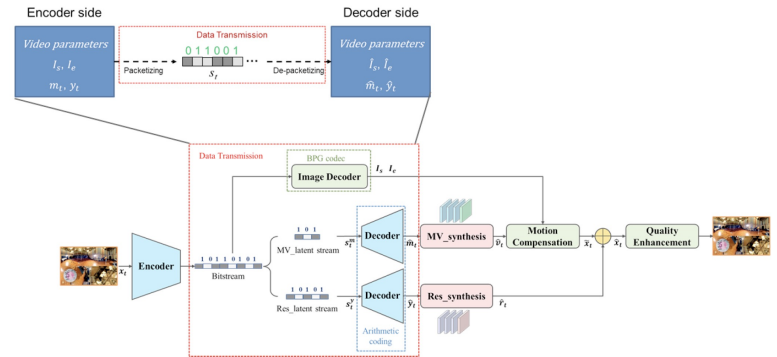
- End-to-end Technologies
 - Small number of proposals on end-to-end video compression
 - General approach
 - Compress intra frames using existing standards
 - Synthesize motion information and texture
 - Informed by low resolution version of motion compensated frame
 - Multi-frame resolution enhancement
 - Multi-frame texture transfer

or

- Auto-encoder framework



JVET-U0096




JVET-X0043
(Omnidirectional Video)

NN Environments

Environments

- NN Environments
 - JVET does not have an official development environment
 - Participants commonly use a mixture of TensorFlow and PyTorch
 - JVET is also studying a proposal for a “Small Ad-hoc Deep Learning Library”
 - Asserted to have few dependencies and compatible with BSD-3 licensing clauses
 - Available in JVET-W0181

 PyTorch
TensorFlow

Small Ad-hoc Deep Learning Library (SADL)

Language	Pure C++, header only.
Footprint	~3200 LOC, library ~200kB, no dependency
Optimization	Some SIMD at hot spots (best effort)
Compatibility	TF 1.x, 2.x, PyTorch converters
Layer Supports	constants, add, maxPool, matMul, reshape, ReLU, conv2D, mul, concat, max, leakyReLU
Type support	float, int32, int16, int8
Quantization	Support adaptive quantizer per layer

JVET-W0181

Conclusions and Additional Info

Conclusions

- Conclusions
 - Overview of technologies considered in JVET's neural network-based video coding exploration
 - Activity includes enhancement of existing video coding tools, introduction of post-filters, and end-to-end video coding system
 - Significant development in neural network-based filtering
 - In-loop filtering, post-filter, and super-resolution
 - Observed gains of 13% in random access

Additional Information

- More information
 - All documents available at <https://jvet-experts.org/>
 - Useful pointers

	21 st Meeting	22 nd Meeting	23 rd Meeting	24 th Meeting
AHG Report	JVET-U0011	JVET-V0011	JVET-W0011	JVET-X0011
Exploration Experiment Report	JVET-U0023	JVET-V0023	JVET-W0023	JVET-X0023
Results Summary	-	-	JVET-W0182	JVET-X0188



SHARP[®]
LABORATORIES
OF AMERICA