

EE1-1.1: Complexity Reduction on Neural-Network Loop Filter

JVET-AD0156

(Training and inference cross-check of JVET-AC0106)



Overview of EE1-1.1 low complexity NN loop filters

- **EE-1.1 test is based on two low complexity NN loop filters studied in JVET-AC0106**
 - EE1-1.1.1 : CP Decomposition and Fusion (CPDF) filter architecture
 - EE1-1.1.2 : Split Luma-Chroma CPDF architecture (crosscheck if time permits)
 - Additional coding gains of 0.2% for luma and >1% for chroma due to training improvements over JVET-AC0106
- **Test is focused on training and inference crosscheck**
 - Training crosscheck: TensorFlow
 - Inference crosscheck: SADL float and SADL int16
 - Crosscheck report: JVET-AD0250 (Ericsson)
- **EE1-1.1.1 Fused CPDF results**
 - Complexity: 16.2 KMAC/pixel
 - AI gains over NNVC-4.0 anchor (Y, Cb, Cr): -4.66%, -5.86%, -5.94%
 - RA gains over NNVC-4.0 anchor (Y, Cb, Cr): -5.28%, -7.60%, -6.70%

EE1-1.1.1 CP-decomposition and fusion architecture

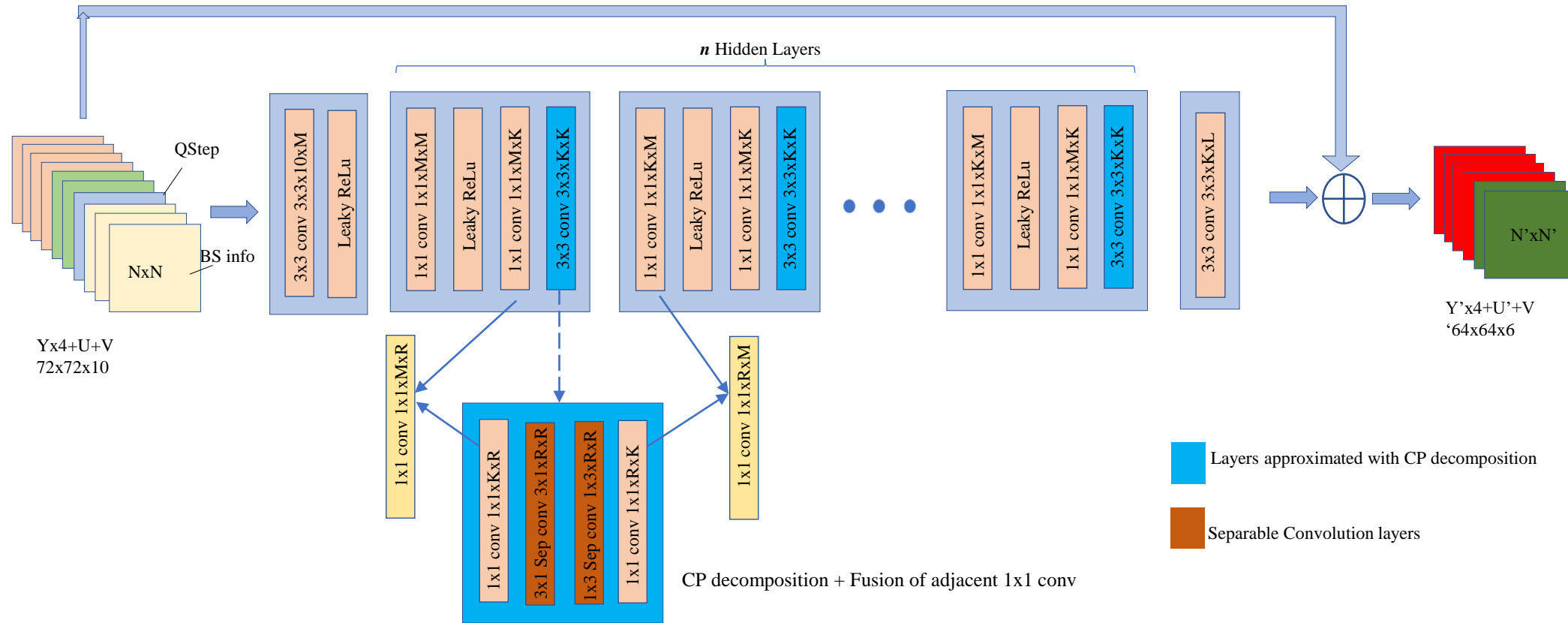


Figure 1 3x3 conv layer CP Decomposition + Fusion of 1x1 conv layers of JVET-X0140 Baseline Model

3x3 convolutions of each hidden layer are decomposed into 4 layers with rank R:

- 1st layer: 1x1xKxR pointwise convolution
- 2nd layer: 3x1xRxR separable convolution
- 3rd layer: 1x3xRxR separable convolution
- 4th layer: 1x1xRxK pointwise convolution

Fusion of adjacent 1x1 conv layers

- No ReLu layers before and after decomposed layer
- 1st decomposed layer merged with previous 1x1 conv
- 4th decomposed layer merged with next 1x1 conv
- Only separable convolution layers left after fusion

Model parameters and Complexity

- M = 72, K = 24, R = 24
- n = 11 (number of hidden layers)
- **Model complexity = 16.2 kMAC/pixel**
- 52.5k params per model

EE-1.1 NN loop filter overview

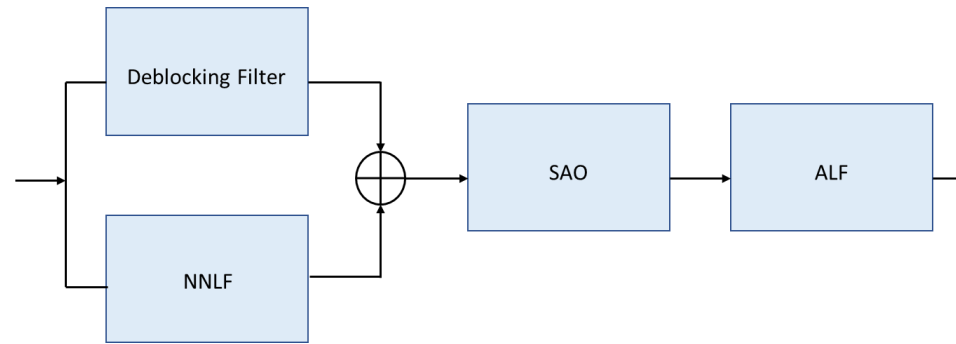
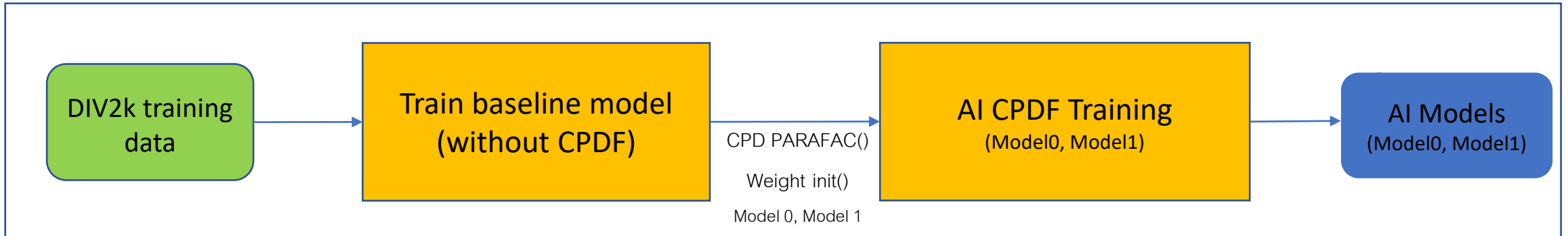


Figure 2 Fusion of deblock and NN filter samples

- NNLF filter placement
 - Reconstructed (pre-deblock) samples used for NN filtering
 - NN Filter applied at 128x128 block level with additional neighbor samples (8 for luma, 4 for chroma) used for filtering
 - NN filter output is combined with deblocked filter using default weights
 - SAO and ALF applied on the fused samples
- Scaling factors at picture level
 - A scaling factor is signaled for each color component in picture headers.
 - Residues are scaled by the scaling factors before addition
- Four models used in filter set
 - Model0 and Model 1: Trained for AI
 - Model2 and Model 3: Trained for RA
 - RDO based model selection done at picture level
 - 128x128 block based RDO decisions for NN filter on and off selection

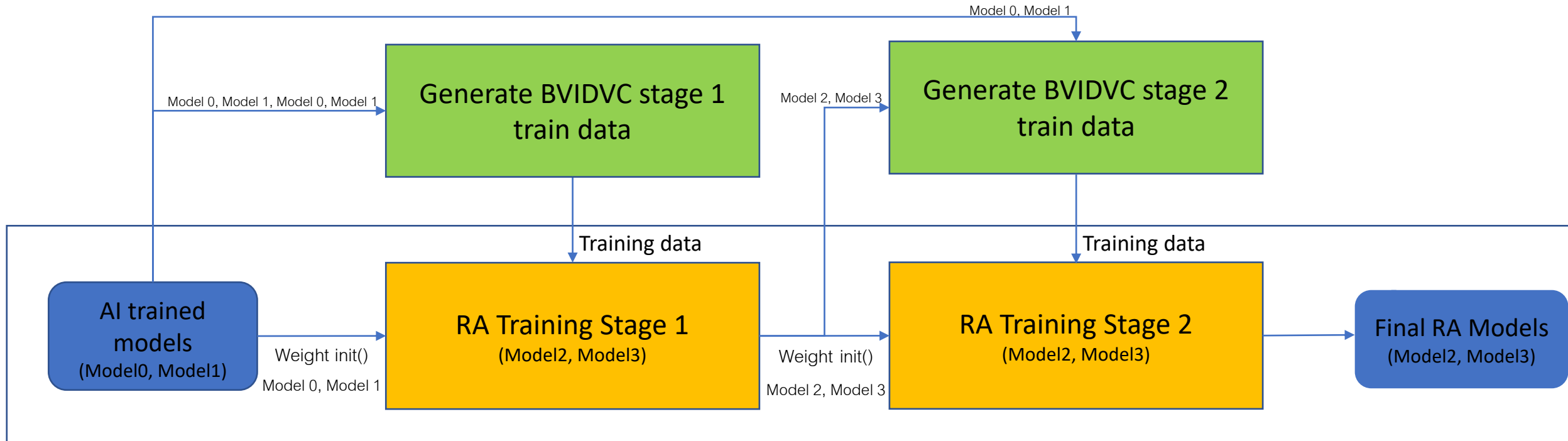
EE-1.1 AI training overview



AI Training overview

- AI training: model 0 and model 1 for I pictures
- TensorFlow version 2.8.0 is used for training
- Generate DIV2k training data for 6 non-CTC QPs
 - Model0 trained using QP 20,30,35 dataset
 - Model1 trained using QP 35,40,45 dataset

EE-1.1 RA training overview



RA Training overview

Training information

Network Information in Training Stage		
Mandatory	GPU Type	Nvidia GeForce RTX3090
	Framework:	Tensorflow 2.8.0
	Number of GPUs per Task	1
	Epoch:	<ul style="list-style-type: none"> • 400 for AI • 40 for RA stage 1 • 30 for RA stage 2
	Batch size:	16
	Training time:	CPD Fused EE-1.1.1: <ul style="list-style-type: none"> - 125 hours for AI - 200 hours for RA stage 1 - 150 hours for RA stage 2
	Training data information:	DIV2K(AI), BVI-DVC(RA)
	Training configurations for generating compressed training data (is different to VTM CTC):	QP=20,25,30,35,40,45
Optional	Number of iterations	
	Patch size	72x72 (4 luma and 2 chroma planes)
	Learning rate:	AI: <ul style="list-style-type: none"> - 2E-4 0-300 epochs - 2E-5 300-350 epochs - 2E-6 350-400 epochs
		RA stage 1: <ul style="list-style-type: none"> - 2E-4 0-20 epochs - 1E-4 20-30 epochs - 2E-5 30-35 epochs - 2E-6 35-40 epochs
		RA stage 2: <ul style="list-style-type: none"> - 2E-4 0-10 epochs - 1E-4 10-20 epochs - 2E-5 20-25 epochs - 2E-6 25-30 epochs
	Optimizer:	ADAM
	Loss function:	Weighted L2 (12,1,1)
	Preprocessing:	Convert 144x144 YUV420 signal to 6 72x72 blocks. Normalize to 0~1

Inference information

- Trained models tested in NNVC-4.0
 - SADL fp32
 - SADL int16
 - Float models renormalized to ensure large weights do not cause overflows
 - SADL quantize() using right shifts modified to enable half rounding

EE1-1.1.1 CPDF Test Results

Table 1 AI performance of EE1-1.1.1 (SADL, fp32)

	All Intra				
	BD-rate Over VTM-11_nnvc-4.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-4.54%	-3.85%	-4.93%	134%	6187%
Class A2	-4.18%	-6.05%	-5.50%	143%	4778%
Class B	-4.14%	-6.45%	-6.30%	127%	4870%
Class C	-4.61%	-7.33%	-7.21%	127%	3621%
Class E	-6.20%	-4.73%	-5.11%	138%	5786%
Overall	-4.66%	-5.86%	-5.94%	132%	4868%
Class D	-5.04%	-5.84%	-6.94%	120%	3743%

Table 2 AI performance of EE1-1.10.1 (SADL, int16)

	All Intra				
	BD-rate Over VTM-11_nnvc-4.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-4.38%	-3.64%	-4.60%	126%	8668%
Class A2	-4.11%	-5.81%	-5.35%	129%	6818%
Class B	-4.09%	-6.17%	-6.02%	119%	6976%
Class C	-4.55%	-7.24%	-7.17%	114%	5235%
Class E	-6.10%	-4.63%	-4.91%	121%	8219%
Overall	-4.58%	-5.67%	-5.74%	121%	6948%
Class D	-4.96%	-5.73%	-6.78%	109%	5388%

Table 3 RA performance of EE1-1.1.1 (SADL, fp32)

	Random Access				
	BD-rate Over VTM-11_nnvc-4.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-5.67%	-4.51%	-4.67%	153%	7327%
Class A2	-5.60%	-6.25%	-4.68%	147%	6427%
Class B	-4.87%	-8.12%	-7.57%	155%	7723%
Class C	-5.24%	-10.27%	-8.63%	139%	7091%
Class E					
Overall	-5.28%	-7.60%	-6.70%	149%	7200%
Class D	-6.47%	-9.31%	-9.18%	138%	7327%

Table 4 RA performance of EE1-1.1.1 (SADL, int16)

	Random Access				
	BD-rate Over VTM-11_nnvc-4.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-5.53%	-4.36%	-4.56%	123%	12761%
Class A2	-5.48%	-6.13%	-4.68%	120%	11395%
Class B	-4.82%	-8.15%	-7.45%	127%	13730%
Class C	-5.21%	-10.16%	-8.51%	124%	12743%
Class E					
Overall	-5.20%	-7.53%	-6.60%	124%	12778%
Class D	-6.43%	-9.12%	-9.13%	128%	13137%

Conclusion

- Salient features of proposed low complexity NN loop filter
 - CP-decomposition
 - Fuse adjacent 1x1 convolutions
- Coding gain vs Complexity
 - Low complexity: 16.2KMAC/pixel
 - 5.28% RA coding gains over NNVC-4.0 anchor
- The contribution proposes to adopt EE-1.1 in NNVC software

THANKS TO ERICSSON FOR CROSS-CHECK