

EE1 related: Performance improvement of AC0052 filter for RPR-based SR in RA configuration

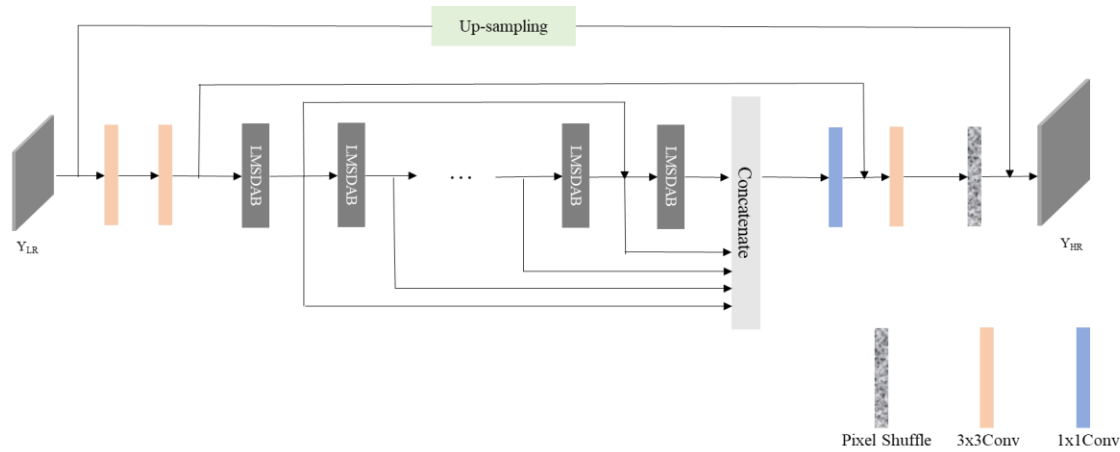
Shimin Huang¹, Cheolkon Jung¹, Yang Liu², Ming Li²

¹Xidian University, China

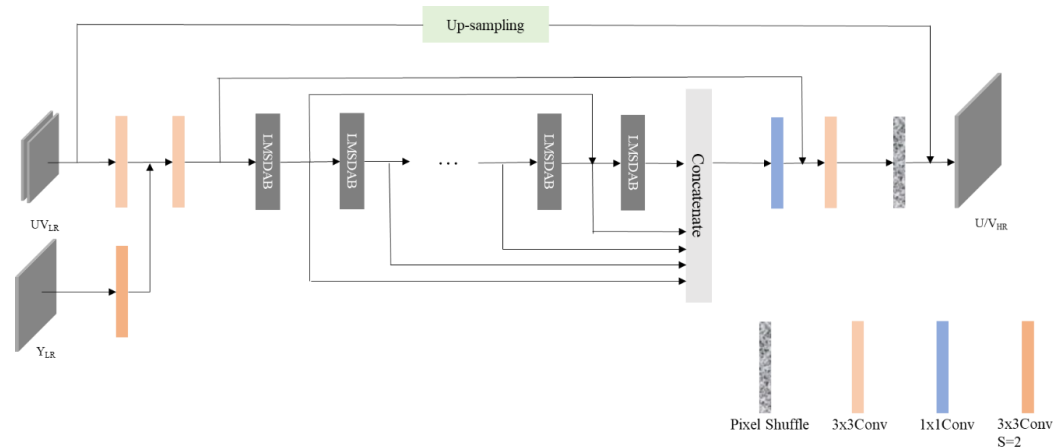
²OPPO, China

Proposed Network

2



Network architecture of the proposed LMSDANet for Y channel.



Network architecture of the proposed LMSDANet for U/V channel.

Proposed Network

3

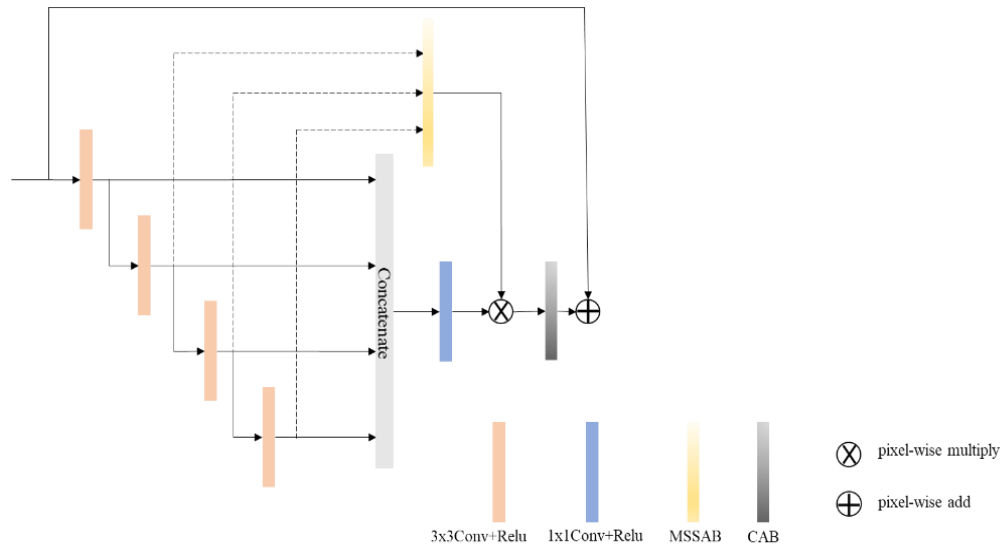


Illustration of the lighter block of multi-mixed scale and depth information with attention (LMSDAB)

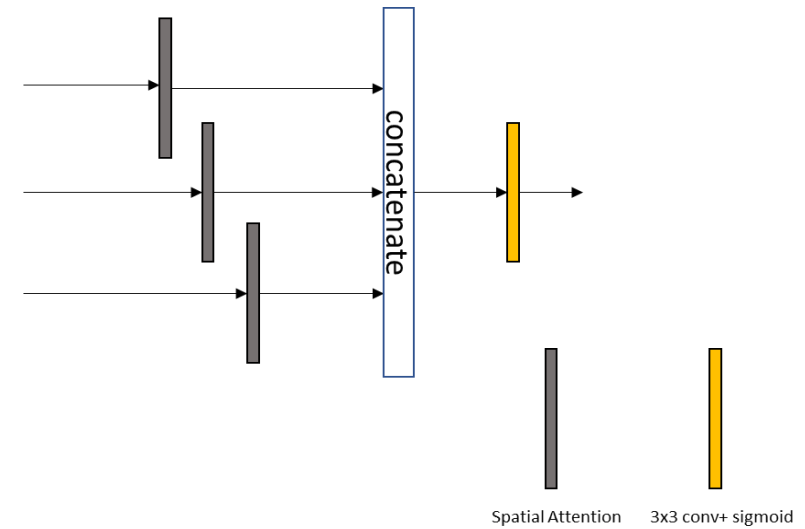
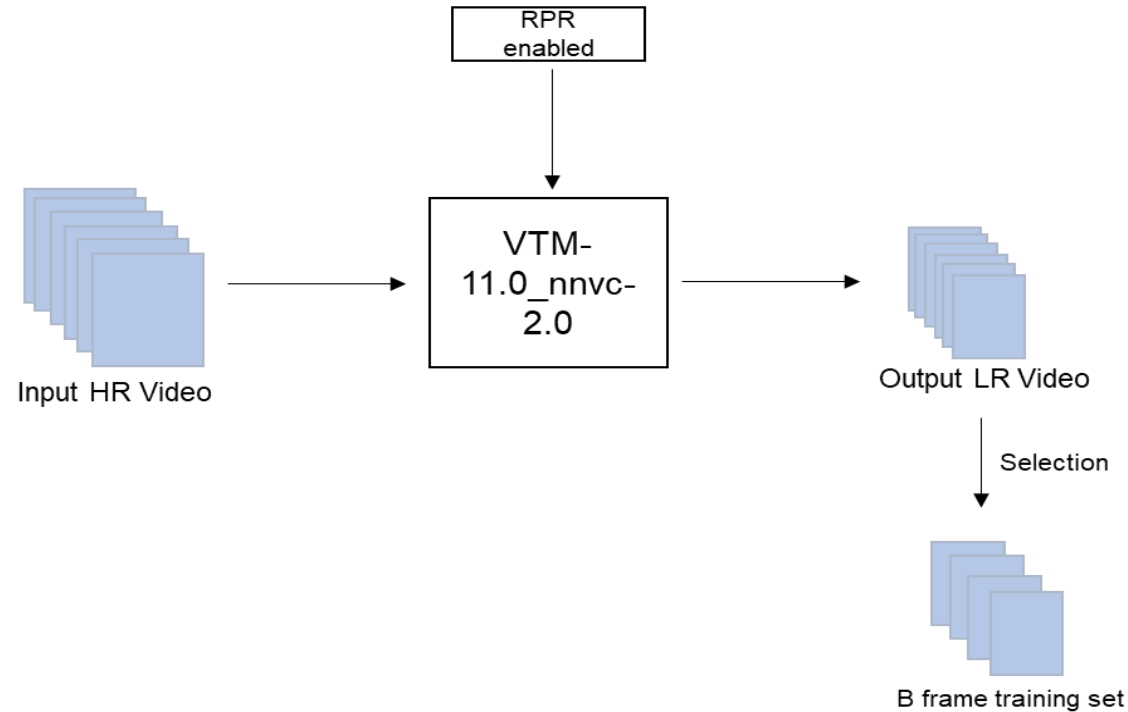


Illustration of the multi-scale spatial attention block (MSSAB)

Generation of B-Frame Model

4

To generate B-frames, we use TVD and BVI-DVC datasets, where we select 200 2K and 200 4K video sequences from BVI-DVC dataset with all video sequences from TVD dataset. These video sequences are compressed by VTM-11.0_NNVC-2.0 in RA configuration with enabled RPR functionality. To form a B-frame training set, some frames from the LR video output are selected as follows:



Process of B-frame training set generation

10% Bit-rate Matching (1)

5

- **JVET-AC0052 filter**

This formula shows that with *BaseQP* as the baseline, if *ScaleF* is increased, then QPs larger than *BaseQP* choose to use the scale factor 2.0x in compression; otherwise, QPs smaller than *BaseQP* choose the scale factor 1.0x. In this formula, *OffsetPSNR* is 0 in RA configuration and 3 in AI configuration. For testing in RA configuration, *BasePSNR* is 42, *BaseQP* is 37, and *ScaleF* is 0.5.

$$RPR_SF[GOP_i] = \begin{cases} 2.0, & DnUpPSNR_i > (BasePSNR - OffsetPSNR) - (QP_i - BaseQP) * ScaleF \\ 1.0, & otherwise \end{cases}$$

$$BaseQP[GOP_i] = BaseQP[GOP_i] - \begin{cases} 5, & RPR_SF[GOP_i] = 2.0 \\ 0, & otherwise \end{cases}$$

10% Bit-rate Matching (2)

6

- **JVET-AC0196 filter**

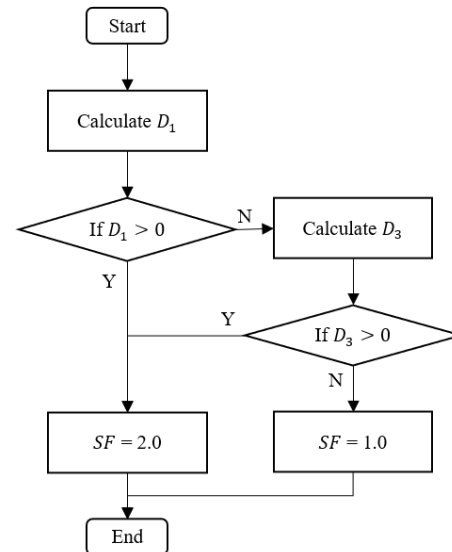
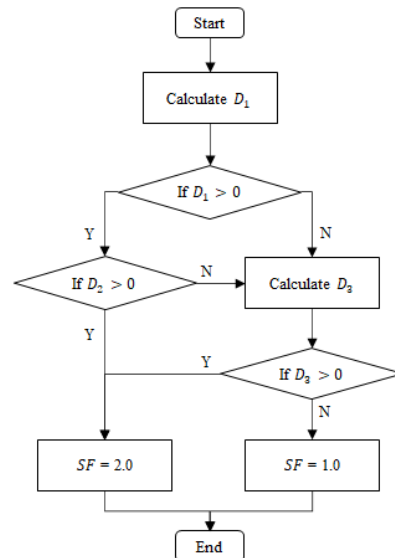
The calculation formulas are as follows:

$$D_1 = DnUpPSNR_{uv} - BasePSNR_{uv}$$

$$D_2 = DnUpPSNR_y - BasePSNR_y$$

$$D_3 = DnUpPSNR_y - (PredefinedPSNR_y - (QP_i - BaseQP) * 0.5)$$

where $DnUpPSNR_y$ and $DnUpPSNR_{uv}$ represent down-up scaled luma and chroma PSNR for the first frame of GOP, respectively. $BasePSNR_y$ and $BasePSNR_{uv}$ represent predefined threshold for luma and chroma PSNR, respectively. $PredefinedPSNR_y$ represents another predefined criteria for luma PSNR and $BaseQP$ represents predefined criteria for QP. QP_i is the initial QP for the first frame of GOP and SF is the determined resampling scale factor.



Network Information in Training Stage

Table 1 Network information for the proposed CNN filter testing in training stage

Network Information in Training Stage		
Mandatory	GPU Type	GPU: NVIDIA 3090 24GB
	Framework:	PyTorch v1.9
	Number of GPUs per Task	1
	Epoch:	luma:50, chroma:50
	Batch size:	32
	Training time:	~26h/model for I, ~50h/model for B
	Training data information:	DIV2K, BVI-DVC & TVD
	Training configurations for generating compressed training data (if different to VTM CTC):	VTM-11.0_NNVC-4.0 , QP {22, 27, 32, 37, 42}
	Loss function:	L2
Optional	Number of iterations	
	Patch size	128×128
	Learning rate:	1e-4
	Optimizer:	ADAM
	Preprocessing:	
	Other information:	

Network Information in Inference Stage

Table 2 Network information for the proposed CNN filter testing in inference stage

Network Information in Inference Stage		
Mandatory	HW environment:	Intel Core i9 12900k @3.9GHz
	Framework:	LibTorch v1.8
	Number of GPUs per Task	0
	Number of Parameters (Each Model)	luma up-sampling model: 1.13M/model chroma up-sampling model: 1.18M/model
	Total Parameter Number	23.15M
	Parameter Precision (Bits)	32 (F)
	Memory Parameter (MB)	20 models in total: 92.6 MB
	Multiply Accumulate (MAC)	964kMAC/pixel
Optional	Total Conv. Layers	91 for up-sampling the luma, 92 for up-sampling the chroma
	Total FC Layers	0
	Batch size:	1
	Patch size	Whole frame

Experimental Results

BD rates in RA configuration over VTM-11.0_NNVC-2.0
with 10% bit-rate matching (1)

Sequences	Random access Main10				
	BD-rate Over NNVC-2.0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-6.61%	-8.24%	-6.42%	70%	#VALUE!
Class A2	-3.50%	2.10%	4.12%	65%	#VALUE!
Class B	0.04%	0.20%	-0.76%	100%	#VALUE!
Class C	0.00%	0.00%	0.00%	100%	#VALUE!
Class E					
Average on A1 and A2	-5.06%	-3.07%	-1.15%	68%	#VALUE!
Overall	-2.01%	-1.16%	-0.71%	85%	#VALUE!

BD rates in RA configuration over VTM-11.0_NNVC-4.0
with 10% bit-rate matching (2)

Sequences	Random access Main10				
	BD-rate Over NNVC-4.0 NnlfOption=0				
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-6.14%	-8.60%	-6.40%	70%	#VALUE!
Class A2	-3.23%	1.21%	3.59%	65%	#VALUE!
Class B	0.00%	0.00%	0.00%	100%	#VALUE!
Class C	0.00%	0.00%	0.00%	100%	#VALUE!
Class E					
Average on A1 and A2	-4.69%	-3.70%	-1.40%	68%	#VALUE!
Overall	-1.88%	-1.48%	-0.58%	85%	#VALUE!

Experimental Results

BD rates of JVET-AC0196 in RA configuration over VTM-11.0_NNVC-3.0 with 10% bit-rate matching

	Random access Main10				
	BD-rate Over NNVC-3.0 NnlfOption=0				
Sequences	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT
Class A1	-7.46%	-8.01%	-7.90%	113%	9700%
Class A2	-3.22%	3.24%	3.73%	123%	1613%
Class B	0.00%	0.00%	0.00%	100%	100%
Class C	0.00%	0.00%	0.00%	100%	100%
Class E					
Average on A1 and A2	-5.34%	-2.38%	-2.08%	118%	5657%
Overall	-2.13%	-0.95%	-0.83%	107%	435%
Class D	0.00%	0.00%	0.00%	100%	100%
Class F	0.00%	0.00%	0.00%	100%	100%

Conclusions

- **BD rate improvement of JVET-AC0052 filter in RA configuration:**
 - Generate the B-frame model using TVD and BVI-DVC datasets
 - Combine two 10% bit-rate matching methods (JVET-AC0052 and JVET-AC0196)
 - Improve the performance of JVET-AC0052 filter in RA configuration by the B-frame model
- **Further study:**
 - Integrate different QP models into a single model
 - Make the proposed CNN filter lighter by depth separable convolution



THANK YOU!

