

EE1-1.5: Ablation Study and Minor Improvements on RTNN

Hao Zhang¹, Cheolkon Jung¹, Yang Liu², and Ming Li²

¹Xidian University, China

²OPPO, China

EE1 Test 1.5

2

This test investigates the NN-based in-loop filtering method proposed in JVET-AC0156. The model explores the combination of residual attention block (RAB) and transformer block (TB) into a loop filter network, and introduces a new attention module to better refine features by introducing auxiliary information.

For EE1 study: possibility to implement in SADL, show gain of each individual element: transformer block, new attention mechanism.

Tests to be studied in EE:

Test EE1-1.5.1: Do the ablation study for transformer block and new attention block.

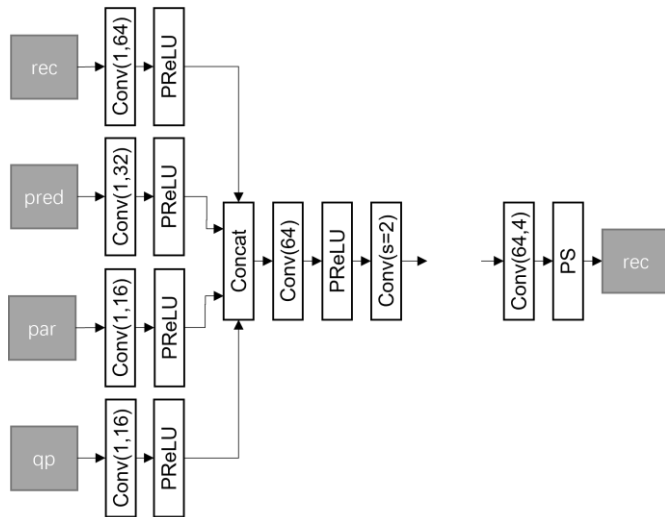
Test EE1-1.5.2: Make minor optimization of the network to further improve performance.

Test EE1-1.5.3: Try to implement the proposed filter using SADL. (We encountered some issues when deploying the model using SADL, so we have contacted crosschecker to replace EE1-1.5.3 with verification of the performance of the JVET-AC0156 model.)

GitLab website: https://vcgit.hhi.fraunhofer.de/jvet-ac-ee1/VVCSoftware_VTM/-/tree/EE1-1.5

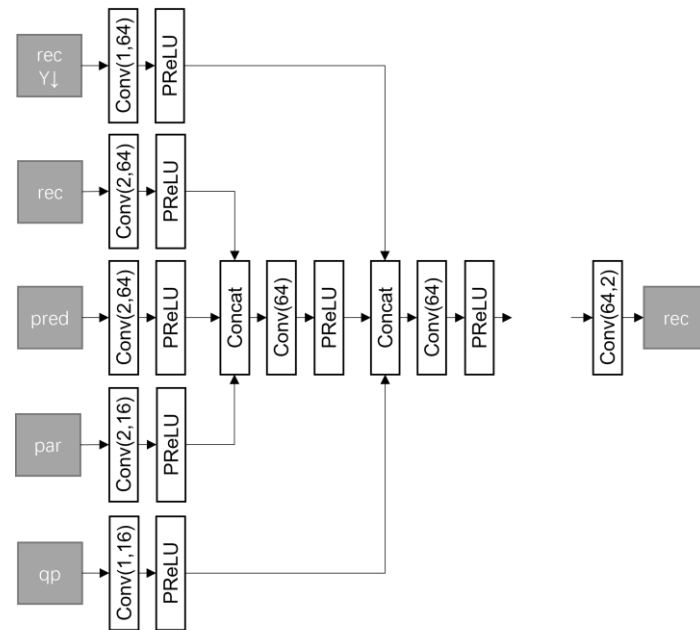
Cross-checker: Dr. M. Santamaria @ Nokia

Proposed Network



(a)

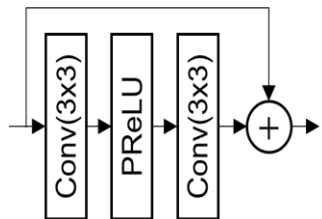
Feature extraction and reconstruction of luma.



(b)

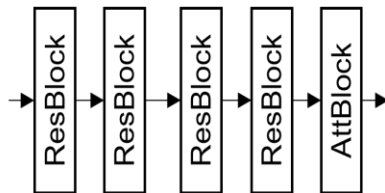
Feature extraction and reconstruction of chroma.

Proposed Network



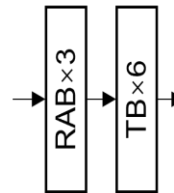
(a)

Residual block.



(b)

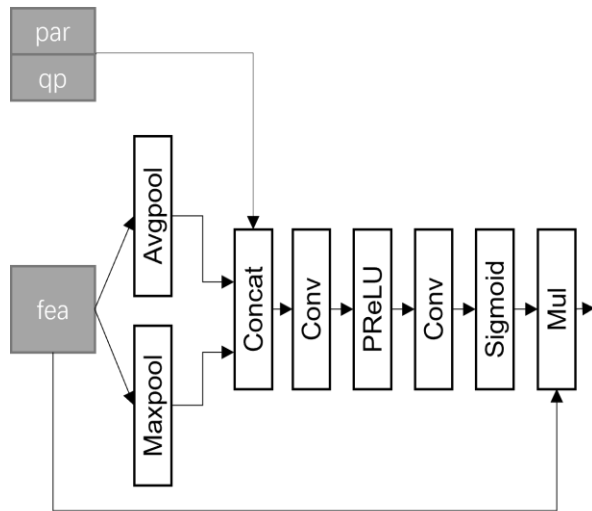
Residual attention block.



(c)

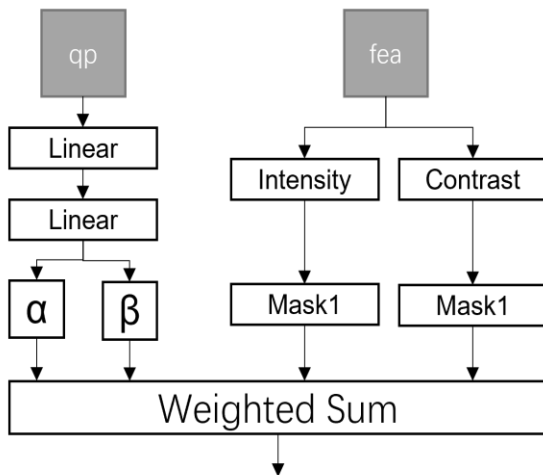
Backbone.

Proposed Network



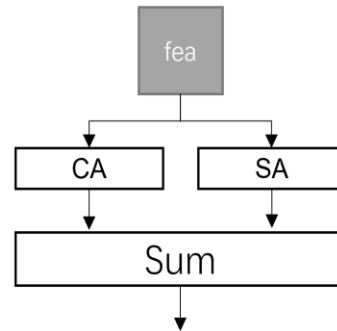
(a)

Spatial attention block.



(b)

Channel attention block.



(c)

Channel-spatial joint attention block.

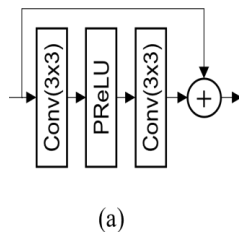
EE1-1.5.2: Minor Improvements

6

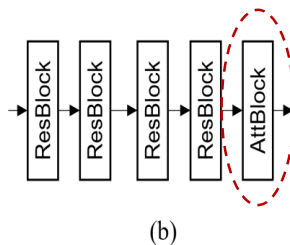
In EE1-1.5.1, we have found that the attention block reduces performance of luma component. However, for chroma component, it improves performance. Thus, we remove the attention block for luma model firstly.

At the same time, we have found that a residual path for communicating the feature of ResBlock and transformer is beneficial for luma component. Because shallow features contain more detailed information, they are important for the recovery of the Y component. Thus, the details of EE1-1.5.2 are as follows:

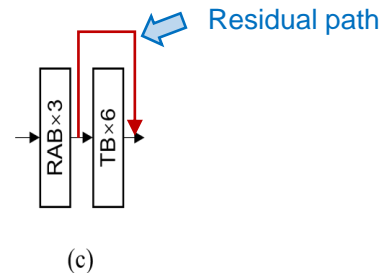
- For **luma model**, the attention block is removed, and a residual path is added.
- For **chroma model**, all the components are not changed.



Residual block (ResBlock).



Residual attention block (RAB).



Backbone.

Network Information in Training Stage

<u>Network Information in Training Stage</u>		
Mandatory	GPU Type	GPU: GeForce RTX 3090
	Framework:	PyTorch v1.9.0
	Number of GPUs per Task	1
	Epoch:	AI (Luma/Chroma): 200 RA(Luma/Chroma): 260
	Batch size:	32
	Training time:	~150h/luma model, ~70h/chroma model
	Training data information:	BVI-DVC, DIV2K
	Training configurations for generating compressed training data (if different to VTM CTC):	VTM-11.0, QP { 12,17,22, 27, 32, 37, 42 }
Optional	Loss function:	L1 and L2
	Number of iterations	
	Patch size	144x144
	Learning rate:	1e-4
	Optimizer:	ADAM
	Preprocessing:	
	Other information:	

Network Information in Inference Stage

Network Information in Inference Stage		
Mandatory	HW environment:	
	GPU Type	CPU only
	Framework:	Libtorch v1.9.0
	Number of GPUs per Task	0
	Number of Parameters (Each Model)	NoTrans: 1.07M (luma), 0.36M (chroma) NoAtt: 1.27M (luma), 0.43M (chroma) All have: 1.4M (luma), 0.47M (chroma) Modified_RTNN: 1.27M (luma), 0.47M (chroma)
	Total Number of Parameters (All Models)	NoTrans: ~2.90M NoAtt: ~3.40M All have: ~3.80M Modified_RTNN: ~3.48M
	Parameter Precision (Bits)	32
	Memory Parameter (MB)	NoTrans: ~11.06M NoAtt: ~12.97M All have: ~14.50M Modified_RTNN: ~13.28M
	Multiply Accumulate (MAC)/pixel	NoTrans: 270K (luma, 144), 358K (chroma, 72) NoAtt: 374K (luma, 144), 448K (chroma, 72) All have: 393K (luma, 144), 482K (chroma, 72) Modified_RTNN: 374K (luma, 144), 482K (chroma, 72)
Optional	Total Conv. Layers	54 Depthwise convolution layers 174 Convolution layers (4 models total)
	Total FC Layers	
	Total Memory (MB)	
	Batch size:	1
	Patch size	144x144 (luma) 72x72 (chroma)
	Changes to network configuration or weights required to generate rate points	
	Peak Memory Usage (Total)	
	Peak Memory Usage (per Model)	
	Border handling	
	Other information:	

Experimental Results

EE1-1.5.1-No_Transformer

All Intra Main10						
BD-rate Over VTM-11.0_nnvc-3.0						
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1	-6.52%	-12.72%	-14.54%	-6.30%	-14.72%	-16.52%
Class A2	-6.89%	-14.00%	-11.69%	-7.29%	-14.47%	-10.61%
Class B	-6.70%	-10.90%	-14.78%	-6.72%	-14.05%	-17.26%
Class C	-7.11%	-13.73%	-16.26%	-7.01%	-16.80%	-18.23%
Class E	-9.89%	-16.42%	-18.03%	-9.90%	-16.57%	-19.82%
Overall	-7.32%	-13.27%	-15.09%	-7.34%	-15.26%	-16.67%
Class D	-7.14%	-13.30%	-16.43%	-6.46%	-16.60%	-18.58%
Class F	-4.72%	-11.02%	-10.64%	-4.90%	-12.96%	-12.95%

Random access Main10						
BD-rate Over VTM-11.0_nnvc-3.0						
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1	-7.96%	-10.64%	-13.19%	-7.87%	-13.15%	-14.77%
Class A2	-8.70%	-15.03%	-11.35%	-7.90%	-14.96%	-10.11%
Class B	-7.67%	-15.83%	-16.50%	-7.18%	-15.85%	-17.00%
Class C	-8.10%	-17.04%	-17.61%	-7.77%	-15.16%	-15.54%
Class E						
Overall	-8.05%	-14.95%	-15.11%	-7.62%	-14.95%	-14.79%
Class D	-9.41%	-17.49%	-18.41%	-6.68%	-16.42%	-15.47%
Class F	-3.68%	-10.04%	-10.20%	-3.89%	-10.75%	-11.58%

Experimental Results

EE1-1.5.1-No_Attention

		All Intra Main10					
		BD-rate Over VTM-11.0_nnvnc-1.0					
		Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1		-6.64%	-19.63%	-22.57%	-6.71%	-22.63%	-24.48%
Class A2		-8.45%	-23.25%	-25.71%	-8.83%	-25.26%	-25.07%
Class B		-8.22%	-22.88%	-20.91%	-8.11%	-24.75%	-23.49%
Class C		-8.84%	-18.59%	-21.34%	-8.57%	-21.68%	-23.41%
Class E		-11.86%	-24.38%	-23.80%	-11.67%	-25.74%	-26.64%
Overall		-8.74%	-21.70%	-22.56%	-8.69%	-23.97%	-24.42%
Class D		-8.25%	-19.59%	-21.50%	-7.36%	-22.97%	-23.89%
Class F		-5.82%	-15.29%	-14.92%	-6.06%	-17.42%	-17.33%

		Random access Main10					
		BD-rate Over VTM-11.0_nnvnc-1.0					
		Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1		-8.92%	-14.99%	-16.76%	-8.67%	-19.50%	-19.72%
Class A2		-10.35%	-22.14%	-24.11%	-9.10%	-22.81%	-23.75%
Class B		-9.41%	-26.48%	-22.57%	-8.19%	-25.83%	-23.52%
Class C		-9.61%	-19.84%	-20.82%	-8.61%	-18.40%	-19.02%
Class E							
Overall		-9.55%	-21.54%	-21.25%	-8.58%	-21.98%	-21.61%
Class D		-10.65%	-22.21%	-22.31%	-6.96%	-21.28%	-19.39%
Class F		-4.54%	-11.89%	-12.04%	-4.67%	-13.38%	-13.50%

Experimental Results

EE1-1.5.1-With_Att_and_Transformer

All Intra Main10						
BD-rate Over VTM-11.0_nnvc-1.0						
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1	-5.65%	-20.81%	-22.90%	-5.45%	-23.92%	-24.69%
Class A2	-7.55%	-25.16%	-27.86%	-8.31%	-27.14%	-27.21%
Class B	-8.29%	-24.53%	-22.50%	-8.18%	-26.63%	-24.97%
Class C	-8.86%	-19.30%	-22.82%	-8.62%	-22.23%	-24.99%
Class E	-11.90%	-24.60%	-25.38%	-11.64%	-26.30%	-28.60%
Overall	-8.46%	-22.86%	-24.01%	-8.42%	-25.23%	-25.91%
Class D	-8.26%	-20.17%	-22.30%	-7.36%	-23.41%	-24.78%
Class F	-5.82%	-16.06%	-15.57%	-5.97%	-18.27%	-18.95%

Random access Main10						
BD-rate Over VTM-11.0_nnvc-3.0						
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1	-8.55%	-15.71%	-16.33%	-8.14%	-20.58%	-18.67%
Class A2	-10.15%	-22.94%	-25.44%	-9.06%	-23.43%	-25.12%
Class B	-9.42%	-27.65%	-23.69%	-8.24%	-27.23%	-24.13%
Class C	-9.64%	-20.55%	-21.47%	-8.64%	-19.13%	-20.11%
Class E						
Overall	-9.45%	-22.43%	-21.97%	-8.49%	-22.98%	-22.16%
Class D	-10.82%	-22.99%	-23.24%	-7.29%	-21.95%	-19.65%
Class F	-4.50%	-12.64%	-12.36%	-4.62%	-14.40%	-14.75%

Experimental Results

EE1-1.5.1-summary

Transformer BD-rate gain:

- {1.29%, 13.39%, 12.64%} in AI configuration
- {1.56%, 10.53%, 9.51%} in RA configuration

Attention BD-rate gain:

- {-0.30%, 1.63%, 2.07%} in AI configuration
- {-0.11%, 1.25%, 1.09%} in RA configuration

It can be seen that:

- Transformer has a significant overall gain in both luma and chroma.
- Attention module is generally not beneficial to the luma component but is beneficial to the chroma component.
- For low-resolution video sequences (such as Class B, C, D, E, F), the attention module is generally beneficial to the luma component. One possible reason for this is that the content of 4K video frames (Class A1 and A2) is relatively uniform within a coding tree unit (CTU), which can cause the attention mechanism to fail.

Experimental Results

EE1-1.5.2-Modified RTNN

		All Intra Main10					
		BD-rate Over VTM-11.0_nnv3.0					
		Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM
Class A1		-7.16%	-20.21%	-23.01%	-7.05%	-23.52%	-24.69%
Class A2		-8.38%	-25.13%	-28.11%	-8.91%	-27.15%	-27.42%
Class B		-8.36%	-24.64%	-22.38%	-8.25%	-26.82%	-24.86%
Class C		-8.92%	-19.62%	-22.67%	-8.68%	-22.73%	-24.86%
Class E		-11.75%	-25.40%	-25.29%	-11.63%	-26.82%	-28.46%
Overall		-8.85%	-22.99%	-23.99%	-8.82%	-25.42%	-25.86%
Class D		-8.30%	-20.67%	-23.03%	-7.45%	-23.65%	-25.44%
Class F		-5.75%	-15.89%	-14.82%	-6.07%	-18.16%	-18.26%

Modified RTNN achieves **0.42% BD-rate gain** on average in Y-PSNR. Among them, the residual path achieves 0.12% BD-rate gain on average for Y channel.

Conclusions

- Perform some ablation studies on RTNN to explore the effects of transformer block and attention block by EE1-1.5.1
- Conduct some minor optimization to further improve RTNN's performance by EE1-1.5.2
- Experimental results show that the transformer has a significant overall gain in both luma and chroma, while the attention module is generally not beneficial to the luma component but is beneficial to the chroma component.



THANK YOU!

