

JVET-AD0189

EE1-RELATED: A SIMPLIFIED NN-BASED RDO MODEL FOR FILTER SET #1

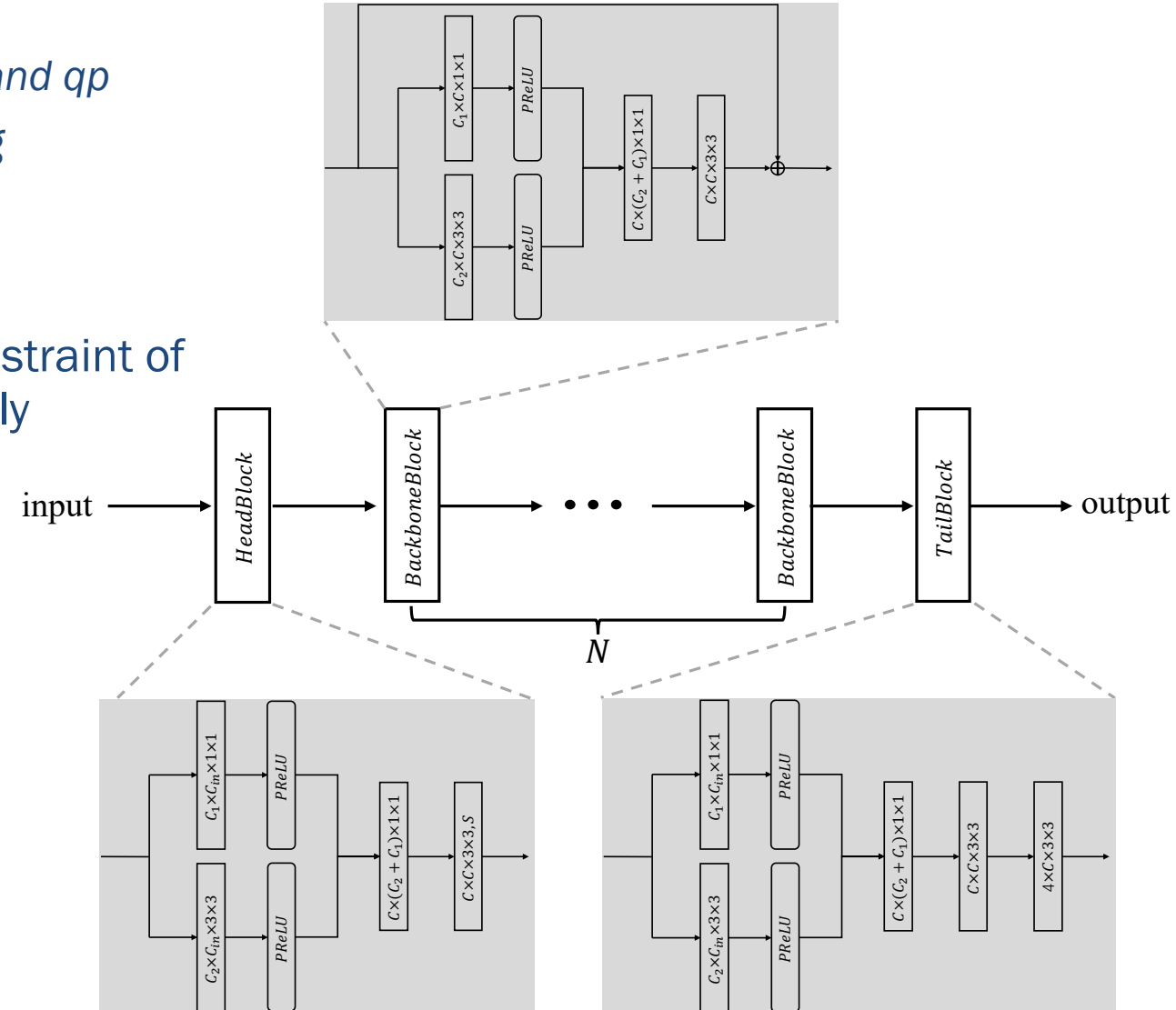
Yue Li, Junru Li, Kai Zhang, Li Zhang

Overview

- A simplification for filter set #1 RDO model to reduce complexity
 - *Unified model for intra and inter*
 - *Much smaller model size*
- Proposed NN-based RDO design
 - *RA: -0.34%, -0.02%, 0.02%, EncT 106%, DecT 100%*
 - *LDB: -0.24%, 0.46%, 0.81%, EncT 107%, DecT 100%*
 - *AI: -0.42%, -0.35%, -0.47%, EncT 114%, DecT 99%*
- NNVC-4.0 NN-based RDO design
 - *RA: -0.31%, -0.10%, -0.16%, EncT 108%, DecT 101%*
 - *LDB: -0.22%, 0.26%, 0.08%, EncT 111%, DecT 100%*
 - *AI: -0.26%, -0.12%, -0.28%, EncT 123%, DecT 103%*
- kMAC/pixel and parameter numbers
 - *Proposed RDO model: 31.2 kMAC/pixel (~5 times smaller), 112.6k parameters (~10 times smaller)*
 - *NNVC-4.0 RDO models: 166.9 kMAC/pixel, 1080.1k parameters*

Proposed method

- Unified model for intra and inter
 - $C_{in} = 4$, i.e. input rec, pred, slice type, and qp
 - $S = 2$ to achieve feature down-sampling
 - $C = 32, C_1 > C, C_2 < C$
 - $N = 4$
- Given the simplified RDO model, the constraint of applying the RDO model is relaxed slightly



Inference

- SADL is used for performing the inference of the proposed CNN filters
- Both weights and internal computations are represented with int16 precision

	kMAC/pixel (frame basis)	Total Parameter Number	Total parameter memory (int16)
Proposed RDO model	31.2	112.6k	225.2k
NNVC-4.0 RDO models	166.9	1080.1k	2160.2k

Experimental results

■ Proposed (much smaller model size)

Random access Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-0.14%	-0.02%	0.00%	105%	99%
Class A2	-0.39%	-0.15%	-0.14%	106%	101%
Class B	-0.36%	-0.07%	0.09%	107%	100%
Class C	-0.43%	0.15%	0.09%	105%	101%
Class E					
Overall	-0.34%	-0.02%	0.02%	106%	100%
Class D	-0.50%	0.26%	0.19%	106%	104%
Class F	-0.23%	0.05%	-0.13%	103%	98%

Low delay B Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1					
Class A2					
Class B	-0.21%	0.51%	1.11%	107%	100%
Class C	-0.24%	0.24%	0.80%	106%	100%
Class E	-0.27%	0.65%	0.30%	107%	101%
Overall	-0.24%	0.46%	0.81%	107%	100%
Class D	-0.35%	0.69%	1.25%	107%	99%
Class F	-0.14%	0.46%	0.42%	104%	102%

All Intra Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-0.21%	-0.28%	-0.21%	112%	100%
Class A2	-0.45%	-0.54%	-0.84%	113%	98%
Class B	-0.40%	-0.43%	-0.76%	112%	97%
Class C	-0.52%	-0.21%	-0.20%	121%	102%
Class E	-0.54%	-0.31%	-0.26%	113%	101%
Overall	-0.42%	-0.35%	-0.47%	114%	99%
Class D	-0.51%	0.08%	-0.03%	114%	100%
Class F	-0.35%	-0.27%	-0.24%	107%	101%

■ Existing

Random access Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-0.14%	-0.10%	-0.01%	105%	101%
Class A2	-0.33%	-0.29%	-0.06%	107%	101%
Class B	-0.33%	-0.01%	-0.15%	110%	103%
Class C	-0.41%	-0.09%	-0.37%	108%	101%
Class E					
Overall	-0.31%	-0.10%	-0.16%	108%	101%
Class D	-0.46%	0.01%	-0.05%	112%	105%
Class F	-0.29%	0.03%	-0.20%	104%	97%

Low delay B Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1					
Class A2					
Class B	-0.19%	0.33%	0.57%	111%	99%
Class C	-0.22%	0.19%	0.22%	114%	101%
Class E	-0.26%	0.25%	-0.94%	108%	101%
Overall	-0.22%	0.26%	0.08%	111%	100%
Class D	-0.26%	-0.24%	0.23%	112%	101%
Class F	-0.13%	0.72%	0.47%	105%	98%

All Intra Main10					
BD-rate Over NNVC-4.0-NnIntra-NnFilterSet1					
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT CPU
Class A1	-0.13%	-0.11%	-0.09%	120%	103%
Class A2	-0.27%	-0.21%	-0.37%	118%	101%
Class B	-0.24%	-0.12%	-0.46%	124%	103%
Class C	-0.33%	-0.04%	-0.28%	129%	106%
Class E	-0.31%	-0.18%	-0.12%	123%	101%
Overall	-0.26%	-0.12%	-0.28%	123%	103%
Class D	-0.31%	-0.06%	0.12%	121%	99%
Class F	-0.19%	-0.26%	-0.20%	111%	101%

Conclusions

- Proposed method (~5 times smaller kmac/pixel, ~10 times smaller model size)
 - RA: -0.34%, -0.02%, 0.02%, *EncT* 106%, *DecT* 100%
 - LDB: -0.24%, 0.46%, 0.81%, *EncT* 107%, *DecT* 100%
 - AI: -0.42%, -0.35%, -0.47%, *EncT* 114%, *DecT* 99%
- Existing method
 - RA: -0.31%, -0.10%, -0.16%, *EncT* 108%, *DecT* 101%
 - LDB: -0.22%, 0.26%, 0.08%, *EncT* 111%, *DecT* 100%
 - AI: -0.26%, -0.12%, -0.28%, *EncT* 123%, *DecT* 103%

This contribution presents a NN-based RDO model with improved performance and reduced complexity

Thank InterDigital for crosschecking!