

JVET-U0115

AHG11: Neural Network-based In-Loop Filter Performance with No Deblocking Filtering Stage

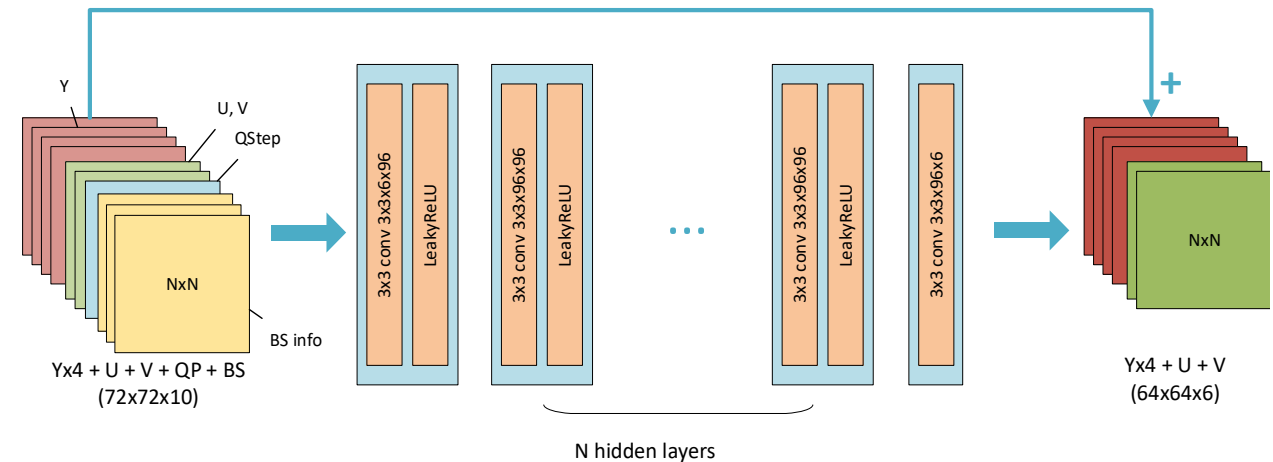
H. Wang, J. Chen, A. M. Kotra, M. Karczewicz (Qualcomm)



Network Structure

Modified on top of subtest 1 of JVET-U0094

- Input: a ~~128x128~~144x144 YUV420 block
 - The input block is converted into a 64x64x6 72x72x6 tensor
 - QP info is used as an additional input plane
 - Deblocking boundary strength (BS) info is used as input planes
 - Pixel values are normalized to [0.0,1.0]
- Output: Residues to be added to input samples
- Number of convolutional layers: 14
- Number of feature maps: 96
- Number of parameters: ~1M



Network Information in Inference Stage			
		JVET-U0094 subtest 1	Proposed method
Mandatory	HW environment:	CPU Only	
	GPU Type	Tensorflow v2.2.0	
	Framework:	0	
	Number of GPUs per Task		
	Total Parameter Number	1,007,814*4Models	1,010,406*4Models
	Parameter Precision (Bits)	32	
	Memory Parameter (MB)	3.84 * 4 models	3.85 * 4 models
Optional	MAC (Giga)	1.57e-4 Giga/pixel	1.99e-4 Giga/pixel
	Total Conv. Layers	14	
	Total FC Layers	0	
	Total Memory (MB)		
	Batch size:	1	
	Patch size	64x64x10	72x72x10
	Peak Memory Usage	1.5MB (input size 128x128)	1.9MB(inputsize 144x144)

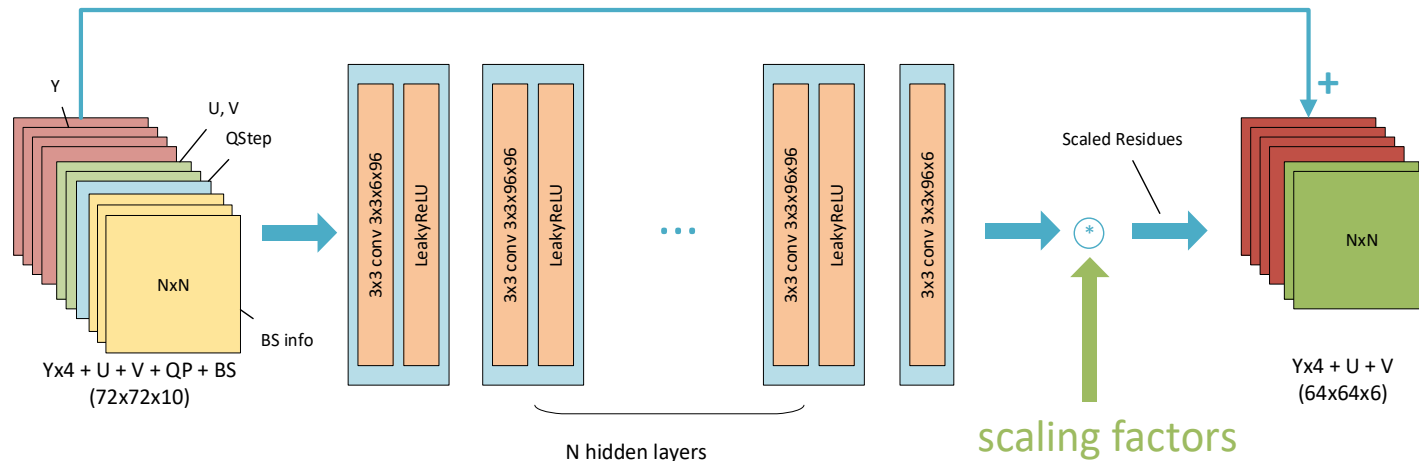
Training process

- Training data set: BVI-DVC
- Training data generation
 - QPs: 20, 25, 30, 35, 40, 45
 - Configurations used for coding BVI-DVC sequences:
 - AI, RA
- Number of trained models:
 - 4 (2AI based + 2RA based)

Network Information in Training Stage		
Mandatory	GPU Type	GPU: NVIDIA Tesla V100-SXM2-32GB
	Framework:	Tensorflow 2.2.0
	Number of GPUs per Task	1
	Epoch:	~130
	Batch size:	13Kx64
	Training time:	89h
	Training data information:	BVI-DVC
Optional	Training configurations for generating compressed training data (if different to VTM CTC):	QP=20,25,30,35,40,45
	Number of iterations	
	Patch size	72x72x10
	Learning rate:	8.00E-04
	Optimizer:	ADAM
	Loss function:	Weighted L2
	Preprocessing:	Convert 144x144 YUV420 signal to 6 72x72 blocks. Normalize to 0~1

VTM integration

- Integrated into VTM-10.0
- For each frame, one model is selected from the 4 candidates
- Slice/CTU level on/off is supported
- A scaling factor is signaled for each color component in the picture header



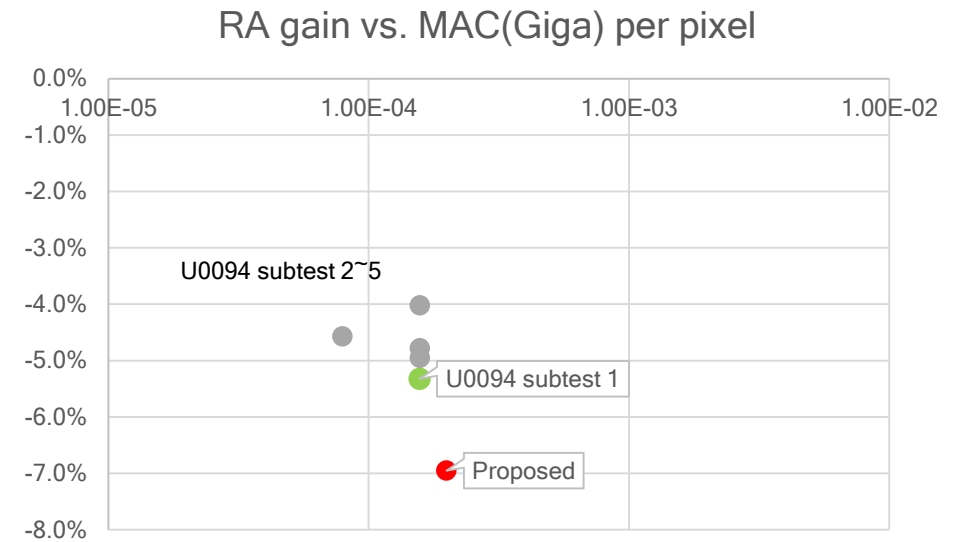
- Deblocking filter module is used to generate BS information but actual filtering is not applied

Simulation results

- QP: 22,27,32,37,42
- Anchor: VTM-10.0 GOP size = 32

Random access Main10						
BD-rate Over VTM-10.0						
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT	bit DIFF
Class A1	-7.10%	-8.89%	-13.28%	145%	13252%	0%
Class A2	-7.46%	-15.28%	-13.78%	140%	11771%	1%
Class B	-6.63%	-14.69%	-15.24%	145%	14161%	0%
Class C	-6.88%	-15.73%	-16.80%	130%	12292%	0%
Class E						
Overall	-6.95%	-13.92%	-14.97%	140%	12969%	0%
Class D	-7.97%	-17.08%	-18.64%	131%	13743%	1%
Class F	-3.48%	-10.73%	-10.16%	170%	16351%	0%
Class H	-6.47%	-9.98%	-13.24%	180%	16645%	0%

All Intra Main10						
BD-rate Over VTM-10.0						
	Y-PSNR	U-PSNR	V-PSNR	EncT	DecT	bit DIFF
Class A1	-5.78%	-9.14%	-12.10%	154%	9600%	0%
Class A2	-5.65%	-12.03%	-10.75%	136%	8291%	0%
Class B	-5.58%	-11.60%	-13.58%	125%	7038%	0%
Class C	-6.05%	-13.31%	-15.69%	118%	5970%	0%
Class E	-7.95%	-12.99%	-14.39%	133%	8634%	0%
Overall	-6.13%	-11.87%	-13.46%	131%	7507%	0%
Class D	-6.27%	-12.75%	-16.56%	117%	6718%	0%
Class F	-4.26%	-10.40%	-10.56%	114%	6859%	0%
Class H	-6.32%	-8.99%	-10.41%	138%	7044%	0%



Thank you!

