

JVET-X0130

AHG11: Cross-component prediction based on a neural network

**Young-Yoon Lee, Tae Meon Bae, Damian Ruiz Coll,
Kalyan Goswami, Alexey Filippov, and Vasily Rufitskiy**

Outline

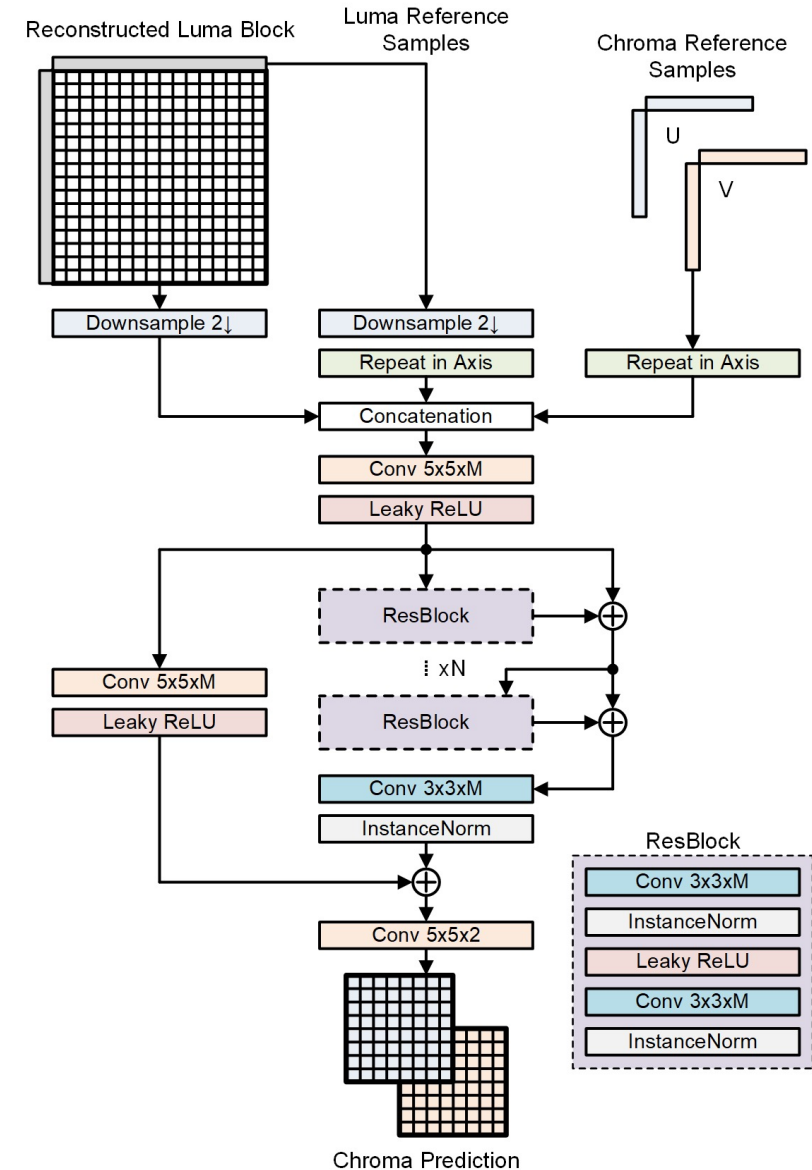
- Introduction
- Proposed method
- Results
- Conclusion

Introduction

- Existing methods
 - The cross-component linear model prediction (CCLM) in VVC
 - The multi-model linear model (MMLM) in ECM
 - Non-linear model based on a convolutional neural network (CNN) model, such as JVET-W0111.
- On top of the cross-component linear model prediction (CCLM), an alternative mode of intra chroma prediction is presented based on a CNN model

Proposed Method

- An input tensor of the size $(W, H, 7)$ is formed from
 - the reconstructed luma block of $(2W, 2H)$,
 - the luma reference samples of $(2W, 1)$ and $(1, 2H)$, and
 - the chroma reference sample of $(W, 1)$ and $(1, H)$.
- The network architecture is based on ResBlock consisting of two 2D convolutional layers:
 - the number of feature maps M is set to 128
 - the number of ResBlocks N is set to 7



Proposed Method

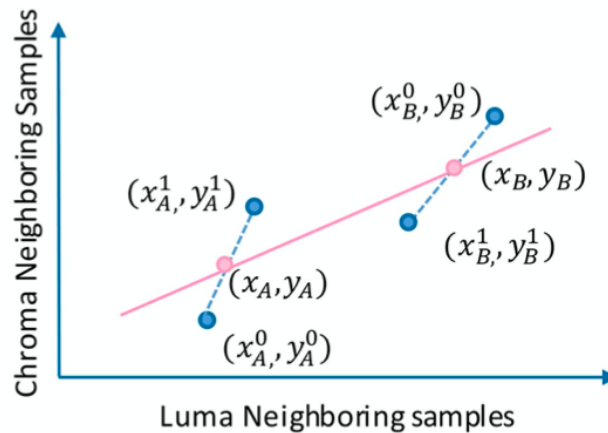
- Network information in training and inference stage

Network Information in Training Stage			Network Information in Inference Stage		
Mandatory	GPU Type	GPU: RTX 3090 x 2 x 24GB	Mandatory	HW environment:	
	Framework:	PyTorch v1.7.1		GPU Type	CPU only
	Number of GPUs per Task	2		Framework:	PyTorch v1.9.0
				Number of GPUs per Task	0
	Epoch:	48			
	Batch size:	512 x 2		Total Parameter Number	2.65M
	Loss function:	L2		Parameter Precision (Bits)	32
	Training time:	37h		Memory Parameter (MB)	10.12
	Training data information:	BVI-DVC		Multiplay Accumulate (MAC) /pixel	2.65M
	Training configurations for generating compressed training data (if different to VTM CTC):				
Optional			Optional		
	Number of iterations			Total Conv. Layers	18
	Patch size	32*32		Total FC Layers	0
	Learning rate:	3.16e-4		Total Memory (MB)	
	Optimizer:	ADAM		Batch size:	1
	Preprocessing:	normalization		Patch size	32x32, 16x16
	Mini-batch selection process:			Changes to network configuration or weights required to generate rate points	
	Other information:			Peak Memory Usage (Total)	
				Peak Memory Usage (per Model)	
				Border handling	
				Other information:	

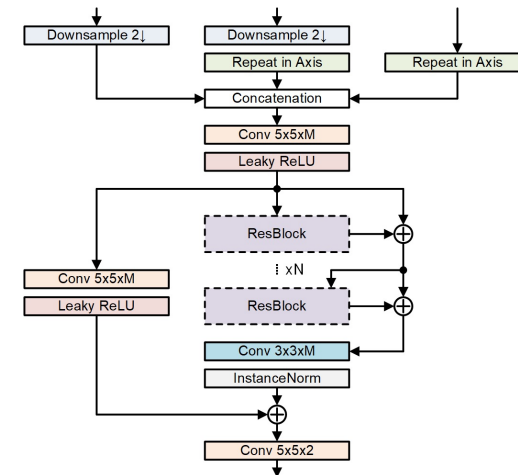
Proposed Method

- Integration with CCLM in VTM
 - For chroma prediction, in addition to three CCLM modes (above-left, above, and left), one more mode is applied for 16x16 and 32x32 chroma blocks.

three CCLM modes



one mode based on CNN



Results

- Test 1: CCNN is enabled for 16x16 and 32x32 chroma blocks.

All Intra Main10								
BD-rate Over VTM-11.0+V0056								
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT
Class A1	-1.19%	0.16%	0.79%	-1.42%	0.38%	0.35%	478%	8442%
Class A2	-1.89%	1.69%	1.66%	-2.21%	1.01%	1.71%	374%	6959%
Class B	-0.54%	0.91%	1.27%	-0.67%	0.70%	1.57%	317%	6397%
Class C	-0.42%	0.88%	1.61%	-0.63%	1.25%	2.76%	272%	5228%
Class E	-0.62%	0.35%	-0.50%	-0.80%	0.96%	0.30%	355%	9782%
Overall	-0.86%	0.82%	1.04%	-1.07%	0.86%	1.44%	344%	6973%
Class D	-0.34%	0.49%	1.89%	-0.58%	1.08%	3.74%	228%	6028%
Class F	-0.30%	0.41%	0.52%	-0.26%	0.99%	1.39%	200%	4637%

Random access Main10								
BD-rate Over VTM-11.0+V0056								
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT
Class A1	-0.38%	-0.37%	0.10%	-0.47%	0.24%	-0.03%	220%	772%
Class A2	-0.72%	0.78%	0.67%	-0.83%	0.61%	1.18%	182%	494%
Class B	-0.23%	0.47%	0.95%	-0.27%	0.31%	1.24%	164%	439%
Class C	-0.10%	0.42%	0.56%	-0.17%	0.21%	1.28%	166%	425%
Class E								
Overall	-0.32%	0.35%	0.62%	-0.40%	0.33%	0.98%	178%	499%
Class D	-0.10%	0.74%	1.11%	-0.23%	0.73%	1.92%	192%	734%
Class F	-0.11%	0.23%	0.03%	-0.09%	0.85%	0.69%	206%	345%

Results

- Additionally, the results of lambda adjustment are provided to deal with the luma-chroma imbalance

- Test 2: CCNN (for 16x16, 32x32 chroma) with lambda adjustment

All Intra Main10								
BD-rate Over VTM-11.0+V0056								
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT
Class A1	-0.56%	-3.71%	-2.44%	-0.66%	-2.83%	-2.62%	493%	9518%
Class A2	-0.90%	-2.55%	-2.17%	-1.19%	-3.26%	-2.32%	383%	7965%
Class B	-0.19%	-3.01%	-2.91%	-0.33%	-3.54%	-3.01%	324%	6900%
Class C	-0.04%	-2.60%	-2.03%	-0.26%	-3.02%	-1.93%	274%	3974%
Class E	-0.36%	-2.83%	-4.02%	-0.54%	-2.56%	-3.68%	360%	9704%
Overall	-0.37%	-2.93%	-2.70%	-0.55%	-3.10%	-2.70%	351%	6982%
Class D	0.03%	-2.93%	-1.76%	-0.18%	-3.30%	-0.96%	231%	4269%
Class F	0.10%	-1.94%	-1.90%	0.00%	-2.24%	-1.65%	203%	3803%

- Test 2a: The anchor with lambda adjustment

All Intra Main10								
BD-rate Over VTM-11.0+V0056								
	Y-PSNR	U-PSNR	V-PSNR	Y-MSIM	U-MSIM	V-MSIM	EncT	DecT
Class A1	0.66%	-4.04%	-3.32%	0.79%	-3.44%	-3.10%	101%	95%
Class A2	1.04%	-4.04%	-3.74%	1.10%	-4.13%	-3.94%	101%	103%
Class B	0.38%	-3.87%	-4.13%	0.37%	-4.38%	-4.60%	101%	102%
Class C	0.37%	-3.64%	-3.61%	0.41%	-4.77%	-4.75%	101%	104%
Class E	0.29%	-3.29%	-3.77%	0.23%	-3.64%	-4.22%	101%	99%
Overall	0.52%	-3.78%	-3.76%	0.55%	-4.14%	-4.21%	101%	101%
Class D	0.37%	-3.43%	-3.54%	0.34%	-4.50%	-4.56%	101%	102%
Class F	0.39%	-2.33%	-2.41%	0.40%	-3.09%	-3.12%	100%	94%

Conclusion

- This contribution presents cross-component prediction model based on a neural network.
- By applying the proposed method, additional BD rate saving can be achieved
- We recommend to study the proposed method in Exploration Experiment (EE) 1.



www.ofinno.com

