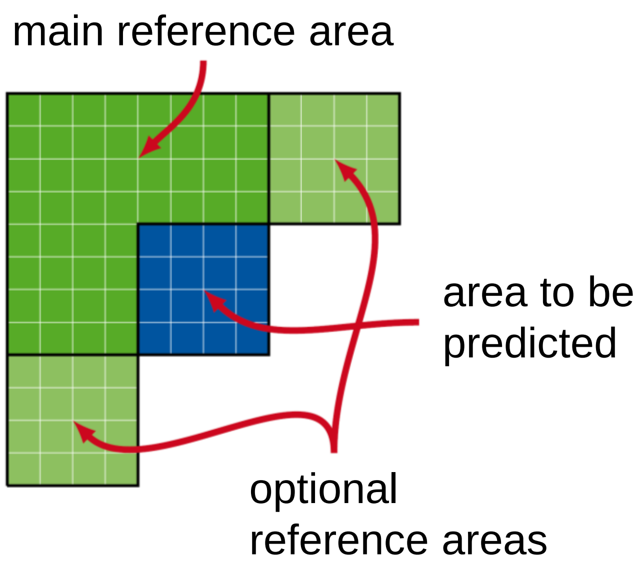
Convolutional Neural Networks for Video Intra Prediction Using Cross-component Adaption

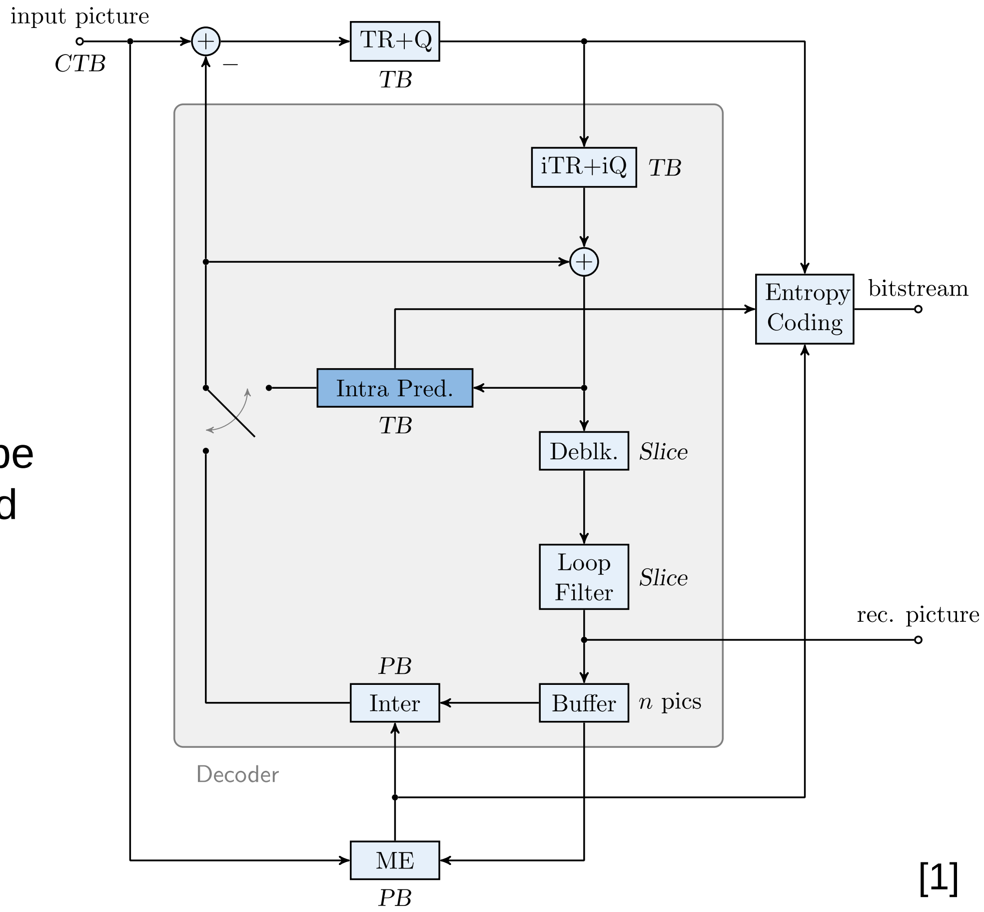
Maria Meyer, Jonathan Wiesner, Jens Schneider, Christian Rohlfing

Problem Statement: Neural Networks for Intra Prediction

Additional neural network (NN) - based intra prediction mode for hybrid video codecs:

- Block-based predictions
- Optionally available information
- Channel wise prediction
- Signaling and rate-distortion decisions
- Low Complexity





[1]

[1] M. Wien, High Efficiency Video Coding – Coding Tools and Specification. Berlin, Heidelberg: Springer, Sept. 2014

Overview

- Open Problems
- Prediction Network
  - Training Methods
  - Architecture
- Mode Signaling and Codec Integration
- Results and Evaluation
- Conclusion

Open Questions

Architecture:

- Best so far can not be definitely concluded due to different training sets
- Only three types of architectures tried so far

Chroma and Cross-Component Prediction:

- No separate consideration of chroma blocks
- No usage of cross component information

Loss Function:

- So far only sum of absolute transform differences (SATD) and mean square error (MSE) compared

Signaling:

- Flag causes a lot of overhead

[2] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, “Fully connected network-based intra prediction for image coding,” IEEE Transactions on Image Processing, vol. 27, no. 7, pp. 3236–3247, July 2018.
[3] Y. Hu, W. Yang, M. Li, and J. Liu, “Progressive spatial recurrent neural network for intra prediction,” Computing Research Repository (CoRR), 2018
[4] T. Dumas, A. Roumy, and C. Guillemot, “Context-adaptive neural network based prediction for image compression,” Computing Research Repository (CoRR), 2018.
[5] J. Pfaff, P. Helle, D. Maniry, S. Kaltenstadler, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand: Neural Network based Intra Prediction for Video Coding, Proceedings of the SPIE 10752, Applications of Digital Image Processing XLI, San Diego, USA, vol. 1075213, September 2018

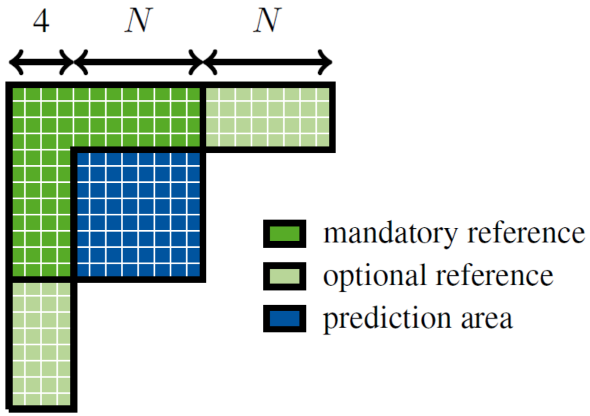
Prediction Network – Luma Architecture

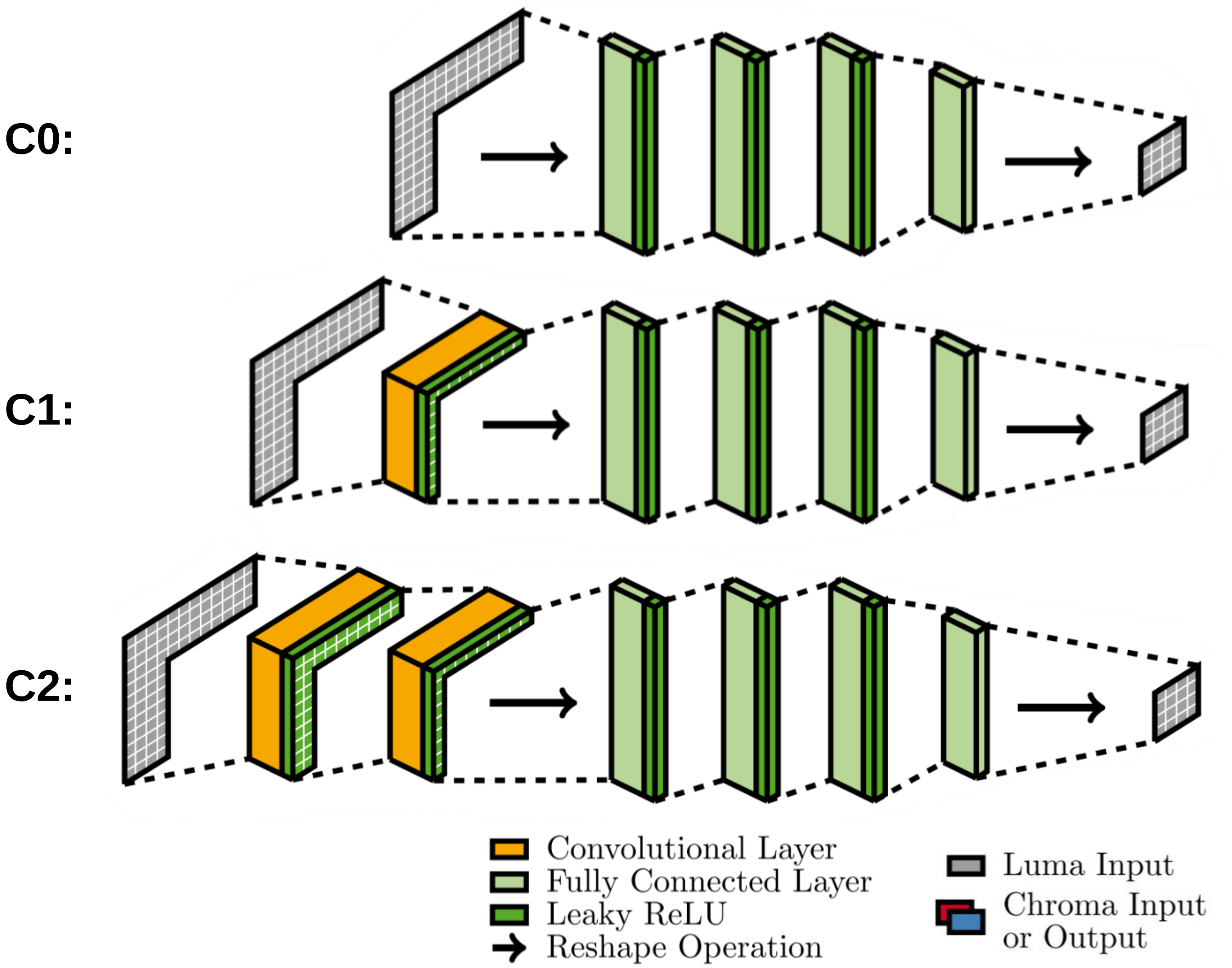
General Settings:

- Four reference lines input
- Separate Networks for each block size

Compared Variants:

- Purely fully-connected architecture (C0)
- Convolutional layers followed by fully-connected ones (C1, C2)





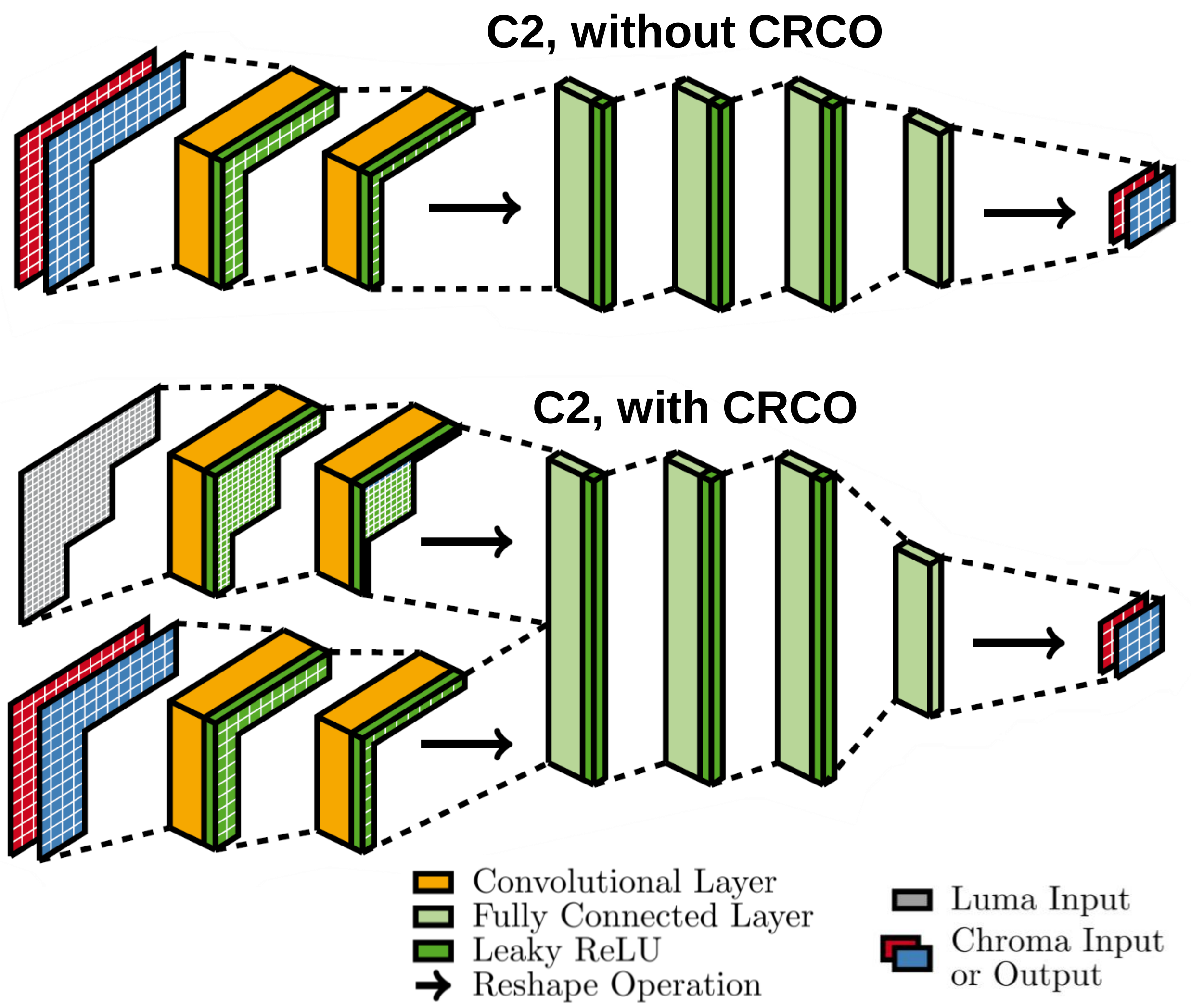
Prediction Network – Chroma Architecture

Joint Chroma Channel Prediction:

- Two input and two output channels
- Otherwise same as luma prediction

Cross-Component Adaptation (CRCO):

- Problems:
  - Different input shape
  - Different resolution
- Architectural Solution:
  - Additional convolutional branch processing luma information
  - Concatenation before first fully connected layer



Prediction Network – Training Methods

Datasets:

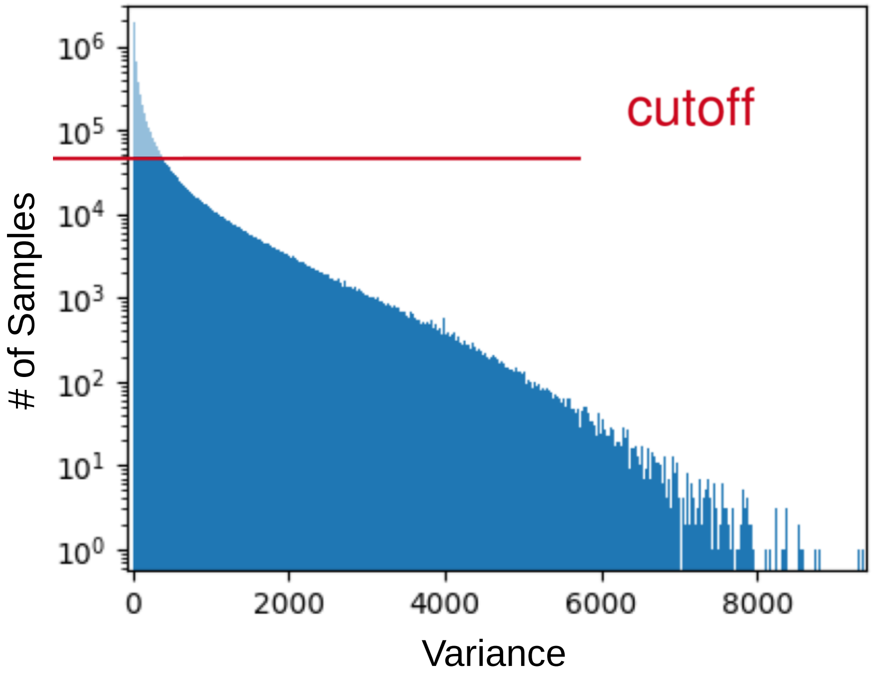
- Extracted samples from 115 raw videos
- Optional input areas masked
- Excluding a portion of the low variance samples possible without loss of bd-rate gains

Training Methods:

- Adam optimizer
- SATD or L1 loss with regularization term

Problems:

- Overfitting for larger chroma blocks



Sample / Parameters Relation

for C2 architecture with CRCO

Blocksize	Luma	Chroma
4	46.29	2.80
8	21.17	0.68
16	7.65	0.12
32	1.10	–

Prediction Examples and Evaluation

Here:

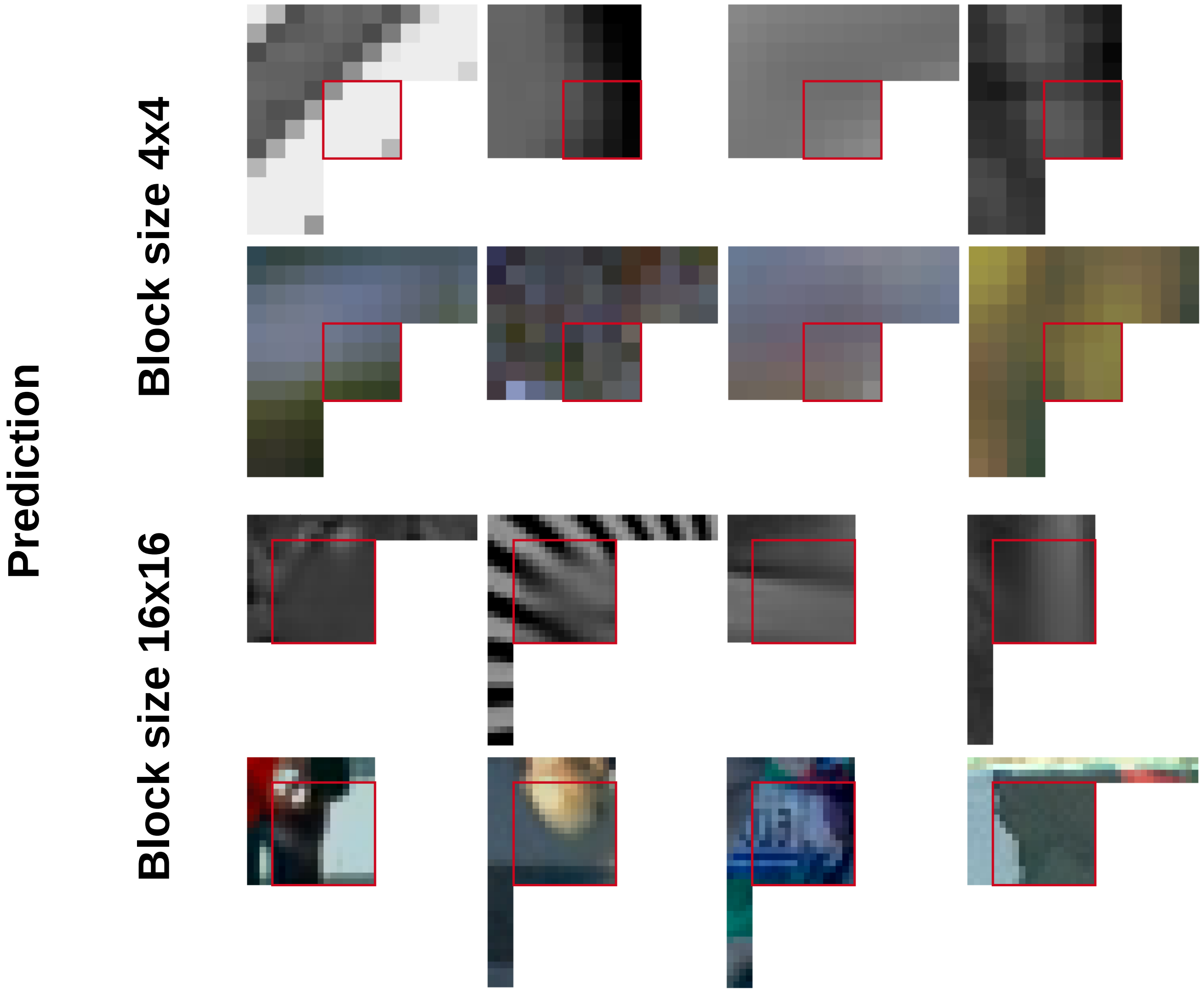
- C2 architecture with CRCO

Luma Samples:

- Enables continuing more than one direction, circles etc.
- Tending towards mean value when continuation unclear/ in bottom right corner

Chroma:

- Enables use of additional luma information



Prediction Examples and Evaluation

Here:

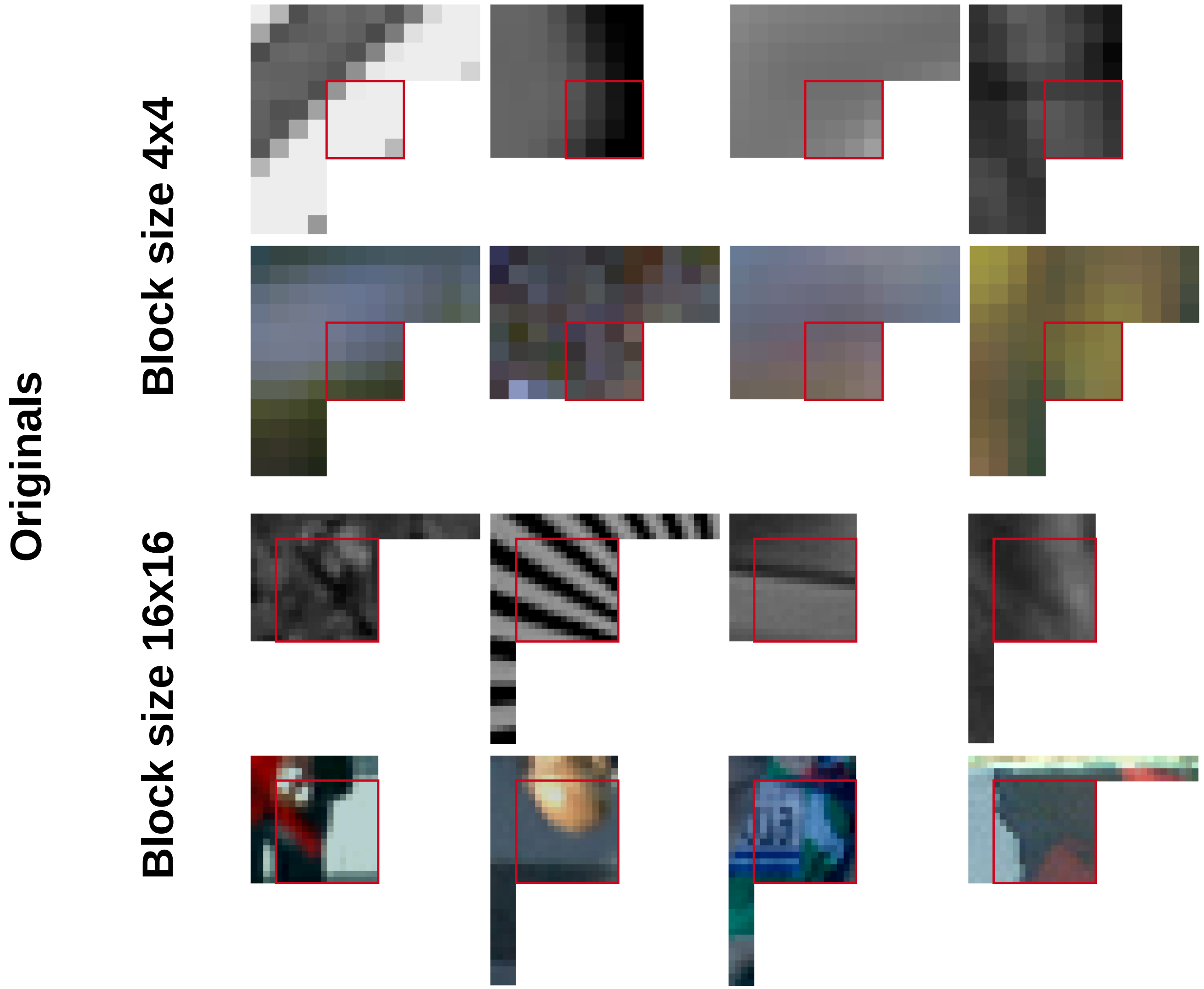
- C2 architecture with CRCO

Luma Samples:

- Enables continuing more than one direction, circles etc.
- Tending towards mean value when continuation unclear/ in bottom right corner

Chroma:

- Enables use of additional luma information



Mode Integration and Signaling

Integration:

- Implemented in HM16.9 as 36th intra mode
- RD-decision as for any other intra mode

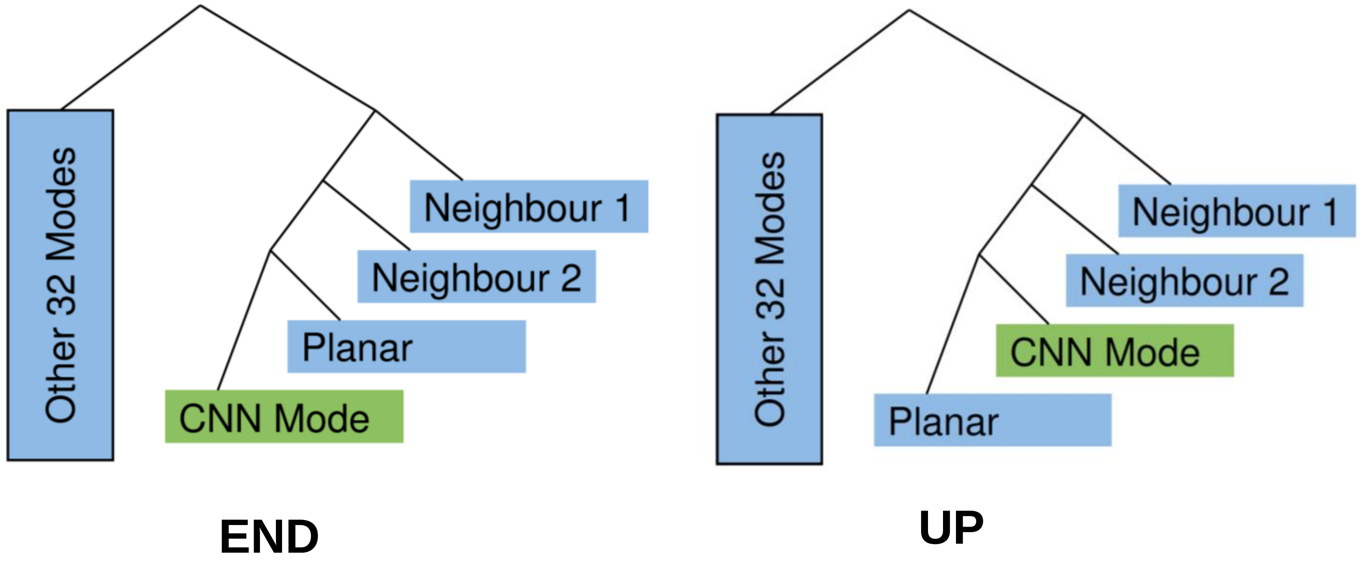
Luma Signaling:

- Most probable mode list extended to four items
- New mode always on MPM-list
- Two variants for MPM-list placement
  - UP: directly behind neighbors
  - END: at the last list position

Chroma Signaling:

- No dedicated signaling for chroma
  - Only useable, when used for luma

Decision Tree Examples



Results – Architecture and Loss

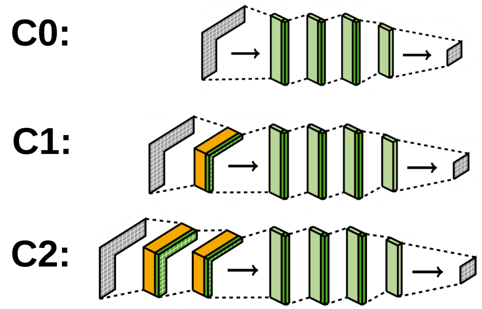
From BD-rates:

- SATD outperforms L1
- C2 outperforms C1 and C0 on average
- C0 better for noisy, high resolution content

Further Analysis:

- C2 always better validation loss
- Difference increasing with block size
- C2 more used for 4x4 blocks, C0 for 32x32 blocks in all class B sequences

Loss Function	L1	SATD
BQTerrace	-1.51 %	**-1.61 %**
BasketballDrive	-1.97 %	**-2.30 %**
Cactus	-2.08 %	**-2.30 %**
Kimono	-2.61 %	**-3.17 %**
ParkScene	-2.75 %	**-2.85 %**
AVG Class B	-2.18 %	**-2.45 %**



Architecture	C2	C1	C0
BQTerrace	**-1.79 %**	-1.74 %	-1.61 %
BasketballDrive	**-2.33 %**	-2.28 %	-2.30 %
Cactus	**-2.46 %**	-2.43 %	-2.30 %
Kimono	-2.66 %	-3.02 %	**-3.17 %**
ParkScene	-2.55 %	-2.66 %	**-2.85 %**
AVG Class B	-2.36 %	-2.43 %	**-2.45 %**
BQMall	**-2.00 %**	-1.85 %	-1.85 %
BasketballDrill	**-1.99 %**	-1.96 %	-1.81 %
PartyScene	**-1.46 %**	-1.39 %	-1.34 %
RaceHorses	**-1.89 %**	-1.84 %	-1.75 %
AVG Class C	**-1.84 %**	-1.76 %	-1.69 %
BQSquare	**-0.98 %**	-0.88 %	-0.79 %
BasketballPass	**-1.85 %**	-1.51 %	-1.49 %
BlowingBubbles	-1.70 %	**-1.74 %**	-1.63 %
RaceHorses	**-2.43 %**	-2.30 %	-2.00 %
AVG Class D	**-1.74 %**	-1.61 %	-1.48 %
AVG All Classes	**-2.01 %**	-1.97 %	-1.91 %

Results – Dedicated Chroma Prediction

Luma Comparison:

- Small improvement (-0.2%) without CRCO
- 3 times more gain (-0.6%) with CRCO

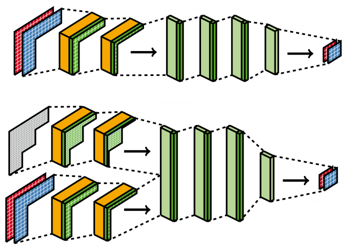
Chroma Comparison:

- Again small improvement (-0.37%) without CRCO
- Nearly -1% with CRCO

Version	with CRCO			without CRCO			no chroma IntraNN		
Channel	**Y**	**U**	**V**	**Y**	**U**	**V**	**Y**	**U**	**V**
BQTerrace	**-1.79%**	**-0.84%**	-0.36%	-1.66%	-0.75%	**-0.79%**	-1.57%	-0.26%	-0.03%
Basket.Drive	**-2.33%**	**-1.64%**	**-1.97%**	-1.83%	-0.15%	-0.88%	-1.34%	-0.05%	-0.56%
Cactus	**-2.46%**	**-1.89%**	**-2.05%**	-1.99%	-1.24%	-1.12%	-1.60%	-0.89%	-0.50%
Kimono	**-2.66%**	**-2.46%**	**-1.84%**	-1.71%	-1.60%	-1.41%	-1.62%	-1.46%	-1.26%
ParkScene	**-2.55%**	**-1.75%**	**-1.91%**	-1.87%	-1.27%	-1.82%	-1.88%	-0.79%	-1.15%
AVG Class B	**-2.36%**	**-1.72%**	**-1.63%**	-1.81%	-1.00%	-1.20%	-1.59%	-0.69%	-0.62%
BQMall	**-2.00%**	**-1.62%**	**-1.57%**	-1.71%	-1.22%	-1.05%	-1.58%	-1.37%	-0.36%
BasketballDrill	**-1.99%**	**-2.42%**	**-2.26%**	-1.21%	-0.38%	-0.74%	-0.63%	-0.17%	-0.21%
PartyScene	**-1.46%**	**-0.86%**	**-0.96%**	-1.31%	-0.68%	-0.70%	-1.21%	-0.68%	-0.65%
RaceHorses	**-1.89%**	**-1.28%**	**-1.44%**	-1.55%	-0.90%	-0.60%	-1.22%	-0.69%	-0.49%
AVG Class C	**-1.84%**	**-1.55%**	**-1.56%**	-1.45%	-0.80%	-0.77%	-1.16%	-0.73%	-0.43%
BQSquare	-0.98%	**-0.65%**	**-0.28%**	**-1.04%**	-0.55%	0.00%	-1.00%	-0.28%	0.20%
BasketballPass	**-1.85%**	**-1.67%**	**-1.36%**	-1.39%	-1.37%	-0.82%	-1.21%	-0.26%	-0.58%
BlowingBubbles	**-1.70%**	**-1.54%**	**-0.97%**	-1.40%	-0.60%	-0.43%	-1.32%	-0.58%	-0.37%
RaceHorses	**-2.43%**	**-1.40%**	**-2.13%**	-1.91%	-1.34%	-1.34%	-1.58%	-0.79%	-0.78%
AVG Class D	**-1.74%**	**-1.32%**	**-1.19%**	-1.44%	-0.97%	-0.65%	-1.28%	-0.48%	-0.38%
AVG All Classes	**-2.01%**	**-1.54%**	**-1.47%**	-1.58%	-0.93%	-0.90%	-1.37%	-0.57%	-0.52 %

Without CRCO:

With CRCO:



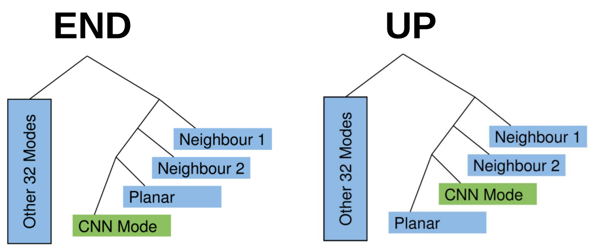
Results – Signaling and Final Evaluation

Signaling:

- UP outperforms end version
  - Mode must be used frequently
- Difference not huge

General Evaluation:

- Hard to compare to other approaches due to training sets
- Beating other approaches in terms of U and V BD-rate gains



Version	**END, with CRCO**			**UP, with CRCO**		
Channel	**Y**	**U**	**V**	**Y**	**U**	**V**
BQTerrace	-1.75%	-0.69%	-0.76%	**-1.79%**	**-0.84%**	-0.36%
Basket.Drive	-2.24%	**-1.64%**	**-2.08%**	**-2.33%**	**-1.64%**	-1.97%
Cactus	-2.35%	**-1.95%**	**-2.06%**	**-2.46%**	-1.89%	-2.05%
Kimono	-2.42%	-2.33%	-1.75%	**-2.66%**	**-2.46%**	**-1.84%**
ParkScene	-2.44%	-1.46%	**-1.99%**	**-2.55%**	**-1.75%**	-1.91%
AVG Class B	-2.24%	-1.61%	**-1.73%**	**-2.36%**	**-1.72%**	-1.63%
BQMall	-1.97%	**-1.63%**	-1.56%	**-2.00%**	-1.62%	**-1.57%**
BasketballDrill	**-2.00%**	-2.17%	-2.03%	-1.99%	**-2.42%**	**-2.26%**
PartyScene	**-1.46%**	-0.83%	-0.95%	**-1.46%**	**-0.86%**	**-0.96%**
RaceHorses	-1.84%	-1.20%	**-1.66%**	**-1.89%**	**-1.28%**	-1.44%
AVG Class C	-1.82%	-1.46%	-1.55%	**-1.84%**	**-1.55%**	**-1.56%**
BQSquare	-1.00%	-0.56%	-0.01%	-0.98%	**-0.65%**	**-0.28%**
BasketballPass	-1.78%	**-1.76%**	-1.19%	**-1.85%**	-1.67%	**-1.36%**
BlowingBubbles	-1.69%	**-1.74%**	-0.71%	**-1.70%**	-1.54%	**-0.97%**
RaceHorses	-2.35%	**-1.92%**	**-2.14%**	**-2.43%**	-1.40%	-2.13%
AVG Class D	-1.71%	**-1.50%**	-1.01%	**-1.74%**	-1.32%	**-1.19%**
AVG All Classes	-1.94%	-1.53%	-1.45%	**-2.01%**	**-1.54%**	**-1.47%**

Conclusion and Outlook

Conclusion:

- Useful to train separate networks for chroma channel prediction and integrate cross component information
- Best Architecture depends on content and complexity restrictions
- SATD loss better approximation than L1
- Proposed new signaling with less overhead

Outlook:

- More architectures, loss functions
- Multiple predictions
- Complexity reduction

Thank you
for your attention

Questions?