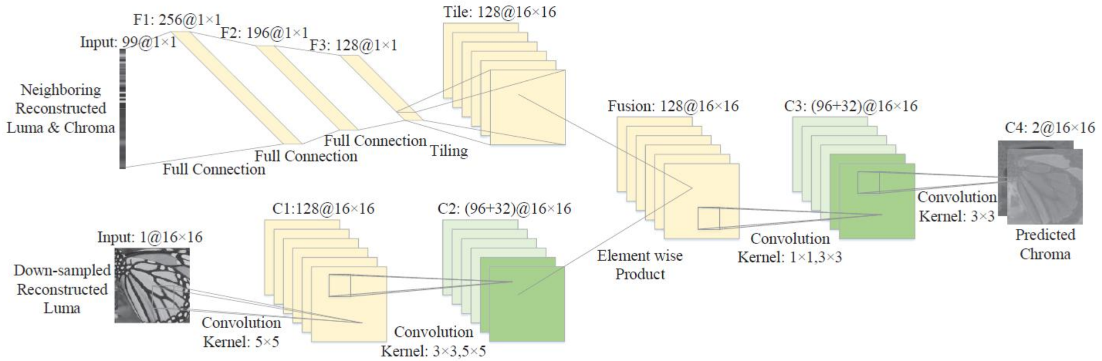
Deep Learning in Compression



Zhu Li
Director, UMKC NSF Center for Big Learning
Dept of Computer Science & Electrical Engineering
University of Missouri, Kansas City
Email: zhu.li@ieee.org, lizhu@umkc.edu
Web: http://l.web.umkc.edu/lizhu

Outline

- A Short Intro do MCC Lab at UMKC
- A Crash Course on Image Coding
- Deep Learning as a compression tool in standard based codec
- End to End Learning based compression
- Summary & Discussions

Short Bio:

Research Interests:

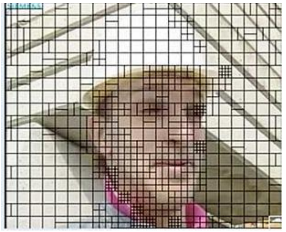
- Immersive visual communicaiton: light field, point cloud and 360 video coding and low latency streaming
- Low Light, Res and Quality Image Understanding
- What DL can do for compression (intra, ibc, sr, inter end2end, c4m)
- What compression can do for DL (model compression, acceleration, feature map compression, distributed training)

NSF I/UCRC Center *for* Big Learning
Creating Intelligence

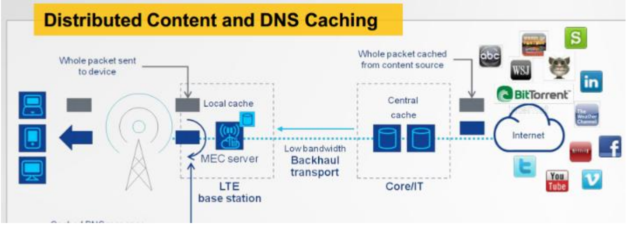
Multimedia Computing & Communication Lab
Univ of Missouri, Kansas City

signal processing and learning

image understanding



visual communication



mobile edge computing & communication

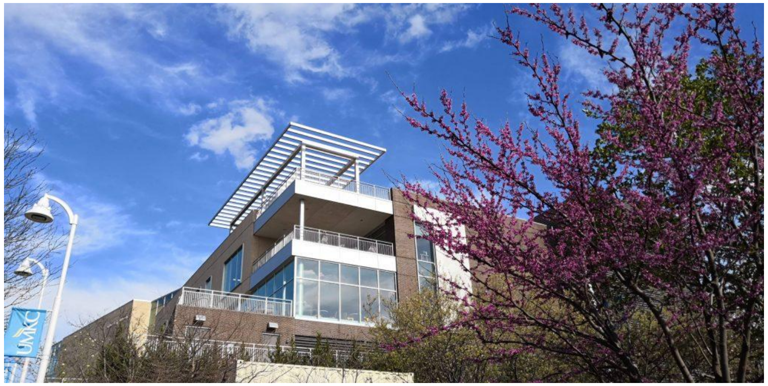
MCC Lab@UMKC

❑ The City

- Sister city of Xian, Edgar Snow's home town

❑ The MCC Lab

- People:
- 2 post-doc, 8 PhDs, 2 visiting PhD on CSC from SJTU and XJTU
- Teaching:
  - Digital Image Processing, Computer Vision, and Video Coding.
- Research focus:
  - Use cases: imaging, compression, and vision
  - Tools: filtering, sparse representation, subspace methods, deep learning, optimization



MCC Lab
http://l.web.umkc.edu/lizhu

- Faculty & Post-Docs:
  - Zhu Li, Northwestern, Lab Director (Fall, 2015~)
  - Li Li, Univ of Science & Tech of China, Visiting Assistant Professor/Asst. Director of MCC Lab (Fall, 2016~)
  - Renlong Hang, Nanjing Univ of Info Science & Tech (NUISCT), Post-Doc Researcher (Fall, 2018~)
- PhD Students
  - Dewan F. Noor, Bangeladesh Univ of Engineering & Tech, PhD Student (Spring, 2016~)
  - Zhaobin Zhang, Huazhong Univ of Science & Tech, PhD Student (Fall, 2016~)
  - Yangfan Sun, MS UMKC, PhD Student (Spring 2017~)
  - Raghunath Puttagunta, MS UMKC, PhD Student (Fall 2017~)
  - Birendra Kathariya, MS UMKC, PhD Student (Fall, 2017~)
  - Anique Akhatar, PhD Student (Spring, 2018~)
  - Wei Jia, B.S and M.S, (Beijing Univ of Post & Telecomm)BUPT, PhD Student (Fall, 2018~)
  - Paras Maharjan, MS/PhD student, Deep learning image enhancement/post processing. 2018~
  - Matthew Kayrish, PhD Student (Spring, 2019~).
  - Han Zhang , visiting PhD student, Shanghai Jiaotong Univ (SJTU), 2018~
  - Wenjie Zhu, visiting PhD Student, SJTU, 2017-18.

Entropy, Conditional Entropy, Mutual Info

- Self Info of an event

$$i(X = x_k) = -\log(\Pr\{X = x_k\}) = -\log(p_k)$$

- Entropy of a source

$$H(X) = \sum_k p_k \log(\frac{1}{p_k})$$

Gaussian: $\ln\left(\sigma\sqrt{2\pi e}\right)$

- Conditional Entropy, Mutual Information

$$H(X_1|X_2) = H(X_1, X_2) - H(X_2)$$

$$I(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2)$$

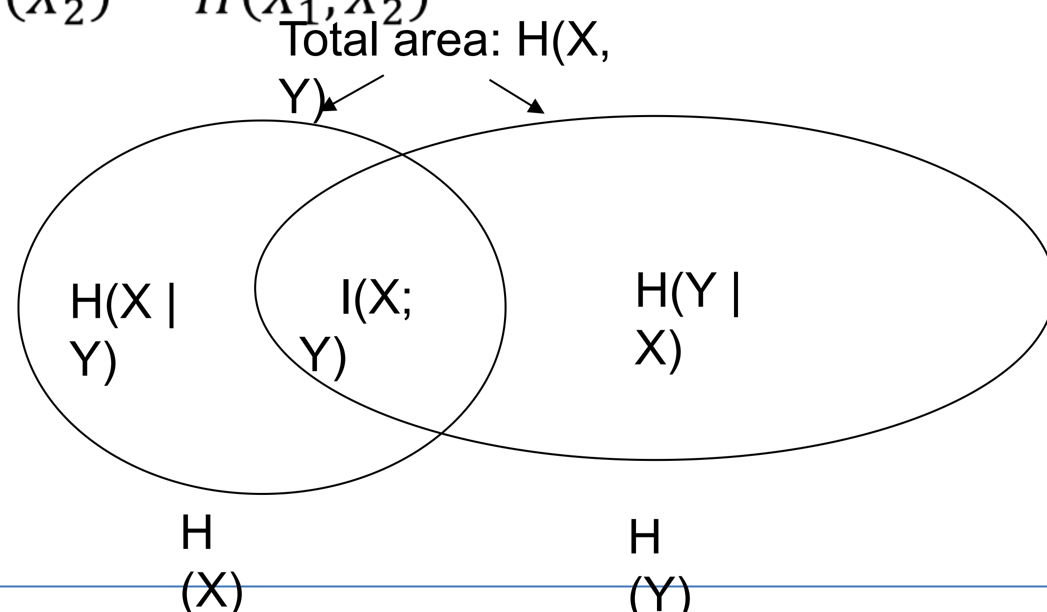
- Relative Entropy

$$D(p||q) = \sum_k p_k \log\left(\frac{p_k}{q_k}\right)$$

Main application: Context Modeling

a b c b c a b

c b a b c b a



Essential of Arithmetic Coding

- Encoding:
  - CDF: $F_x()$=[0.7 0.8 1]

```
LOW=0.0, HIGH=1.0;
while (not EOF) {
    n = ReadSymbol( );
    RANGE = HIGH - LOW;
    HIGH = LOW + RANGE * CDF(n);
    LOW  = LOW + RANGE * CDF(n-1);
}
output LOW;
```

what if PMF not fixed?

Consider a three-letter alphabet $\mathcal{A} = \{a_1, a_2, a_3\}$ with $P(a_1) = 0.7$, $P(a_2) = 0.1$, and $P(a_3) = 0.2$. Using the mapping of Equation (4.1), $F_X(1) = 0.7$, $F_X(2) = 0.8$, and $F_X(3) = 1$. This partitions the unit interval as shown in Figure 4.1.

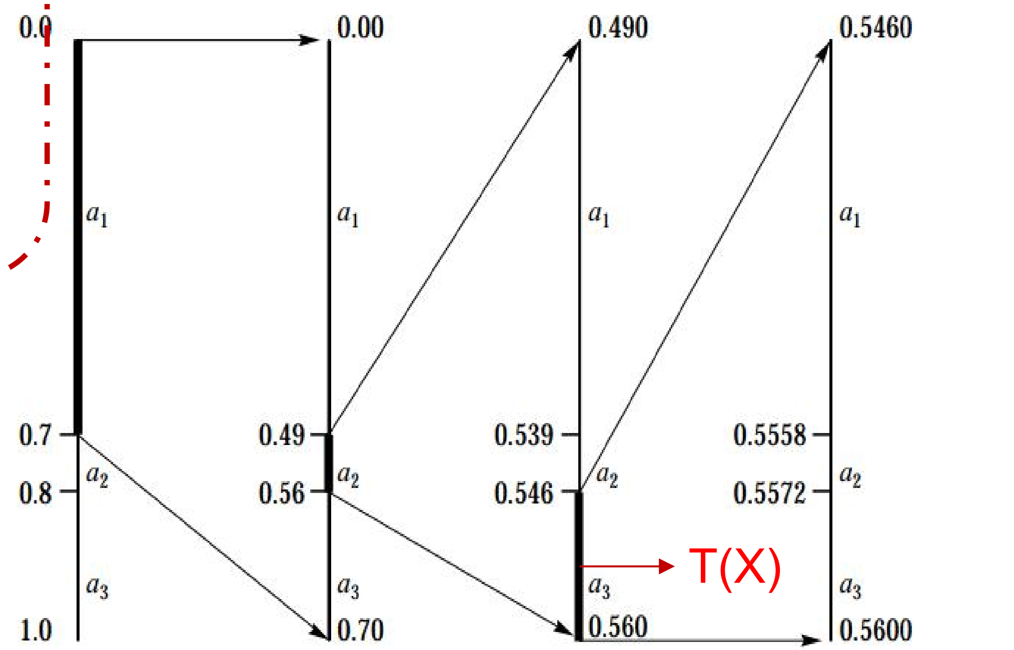


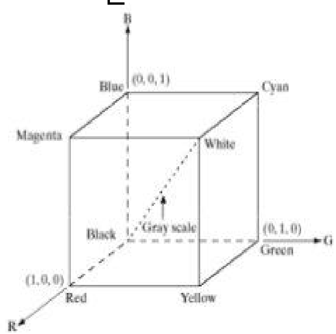
FIGURE 4.1 Restricting the interval containing the tag for the input sequence $\{a_1, a_2, a_3, \ldots\}$.

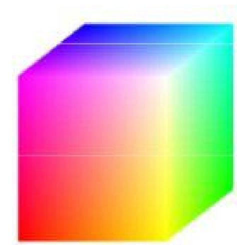
  - T(X)=(0.546+0.560)/2=0.553
  - P(X)= 0.014; l(X)=6 bits
  - Code: $(0.553)_2 = 0.100011011$

YUV/YCbCr/YIQ Model

Rec. 601 for TV: specifies a range of [16, 235] for Y' and [16, 240] for C_B and C_R. To obtain $Y'C_BC_R$ from 8-bit R'G'B' values (i.e., in the range [0, 255]), use the transformation:

$$\begin{bmatrix} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \bullet \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$





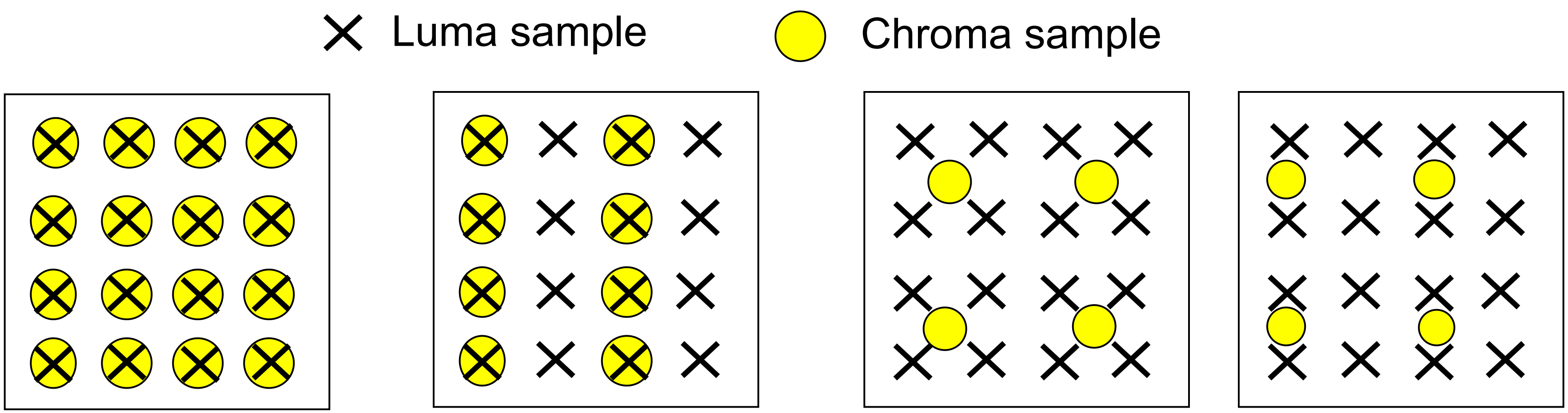
RGB 24-bit color cube

Conversion between RGB and YIQ

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0.956 & 0.621 \\ 1.0 & -0.272 & -0.649 \\ 1.0 & -1.106 & 1.703 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}$$

Color Space Re-Sampling

- RGB components of an image have strong correlation.
  - Can be converted to YUV space for better compression.
- HVS is more sensitive to the details of brightness than color.
- Can down-sample color components to improve compression.



MPEG-1 MPEG-2

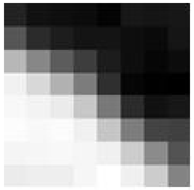
YUV 4:4:4
No downsampling
Of Chroma

YUV 4:2:2
- 2:1 horizontal downsampling of chroma components
- 2 chroma samples for every 4 luma samples

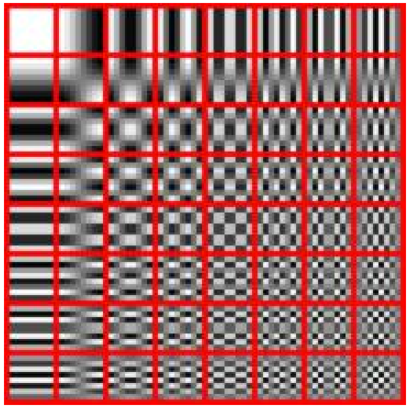
YUV 4:2:0
- 2:1 horizontal downsampling of chroma components
- 1 chroma sample for every 4 luma samples

The Basics of Image Coding

- ❑ Block (8x8 pel) based coding
- ❑ DCT transform to find sparse representation
- ❑ Quantization reflects human visual system
- ❑ Zig-Zag scan to convert 2D to 1D string
- ❑ Run-Level pairs to have even more compact representation
- ❑ Hoffman Coding on Level Category
- ❑ Fixed on the Level with in the category



*



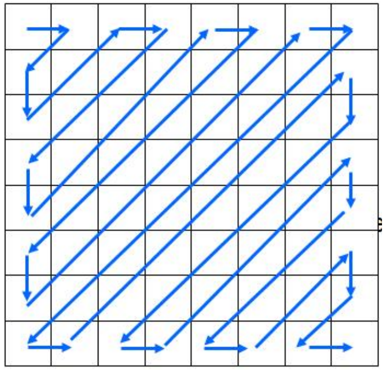
=

Quant Table:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Coding of AC Coefficients

- Zigzag scanning:



- Example

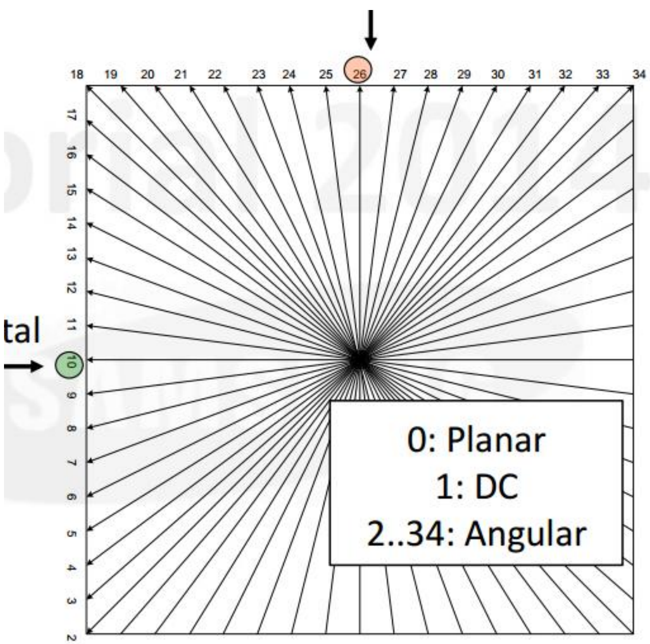
8	24	-2	0	0	0	0	0
-31	-4	6	-1	0	0	0	0
0	-12	-1	2	0	0	0	0
0	0	-2	-1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

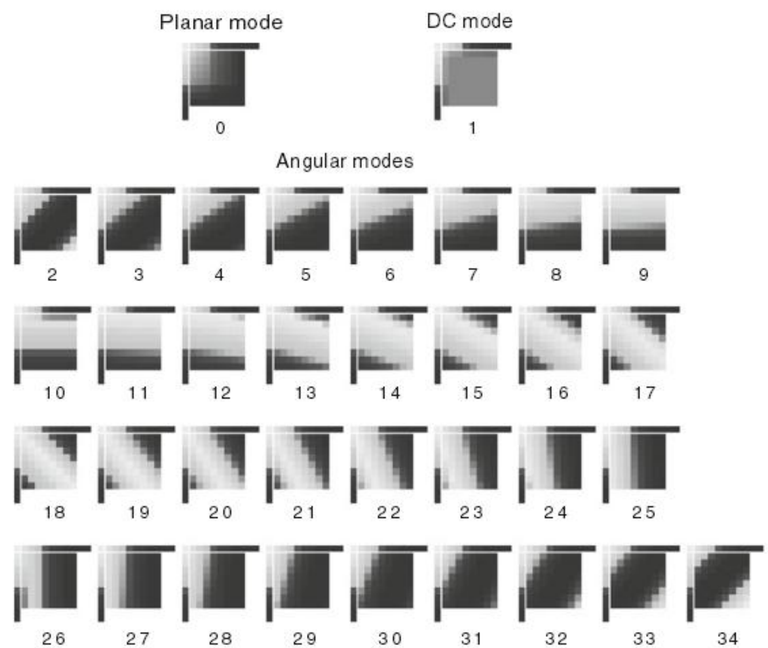
- Example: zigzag scanning result
  24 -31 0 -4 -2 0 6 -12 0 0 0 -1 -1 0 0 0 2 -2 0 0 0 0 0 -1 EOB
- (Run, level) representation:
- (0, 24), (0, -31), (1, -4), (0, -2), (1, 6), (0, -12), (3, -1), (0, -1), (3, 2), (0, -2), (5, -1), EOB

Better Intra Prediction

- Much more modes
  - DC mode: copy DC values from neighbor
  - Planar mode: top row or left col average
  - Angular: pixels on certain line
  - Ref: Jani Lainema, Frank Bossen, Woojin Han, Junghye Min, Kemal Ugur, Intra Coding of the HEVC Standard. *IEEE Trans. Circuits Syst. Video Tech*. 22(12): 1792-1801 (2012)

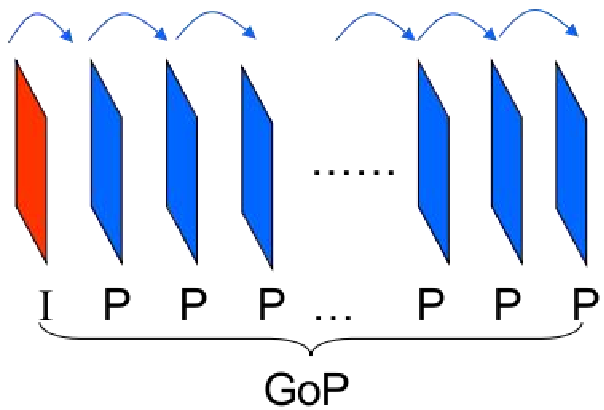
Like a sparse transform basis!

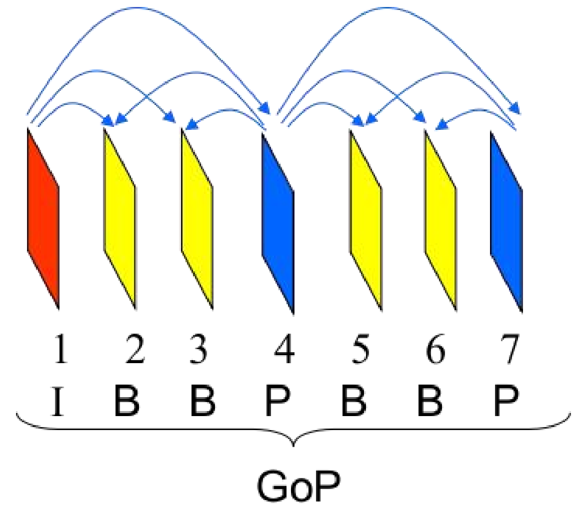




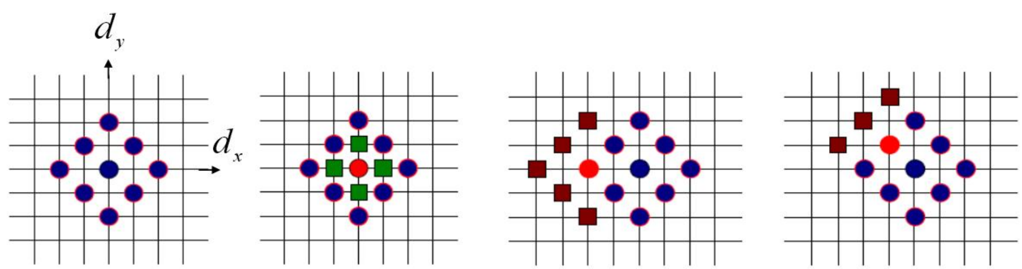
Video Signal Processing

❑ Prediction Structure





- Fast Block Motion Estimation:



Start with large diamond pattern at (0,0)

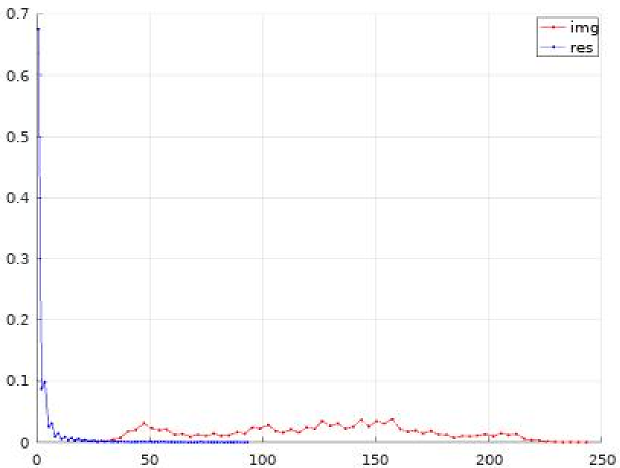
If best match lies in the center of large diamond, proceed with small diamond

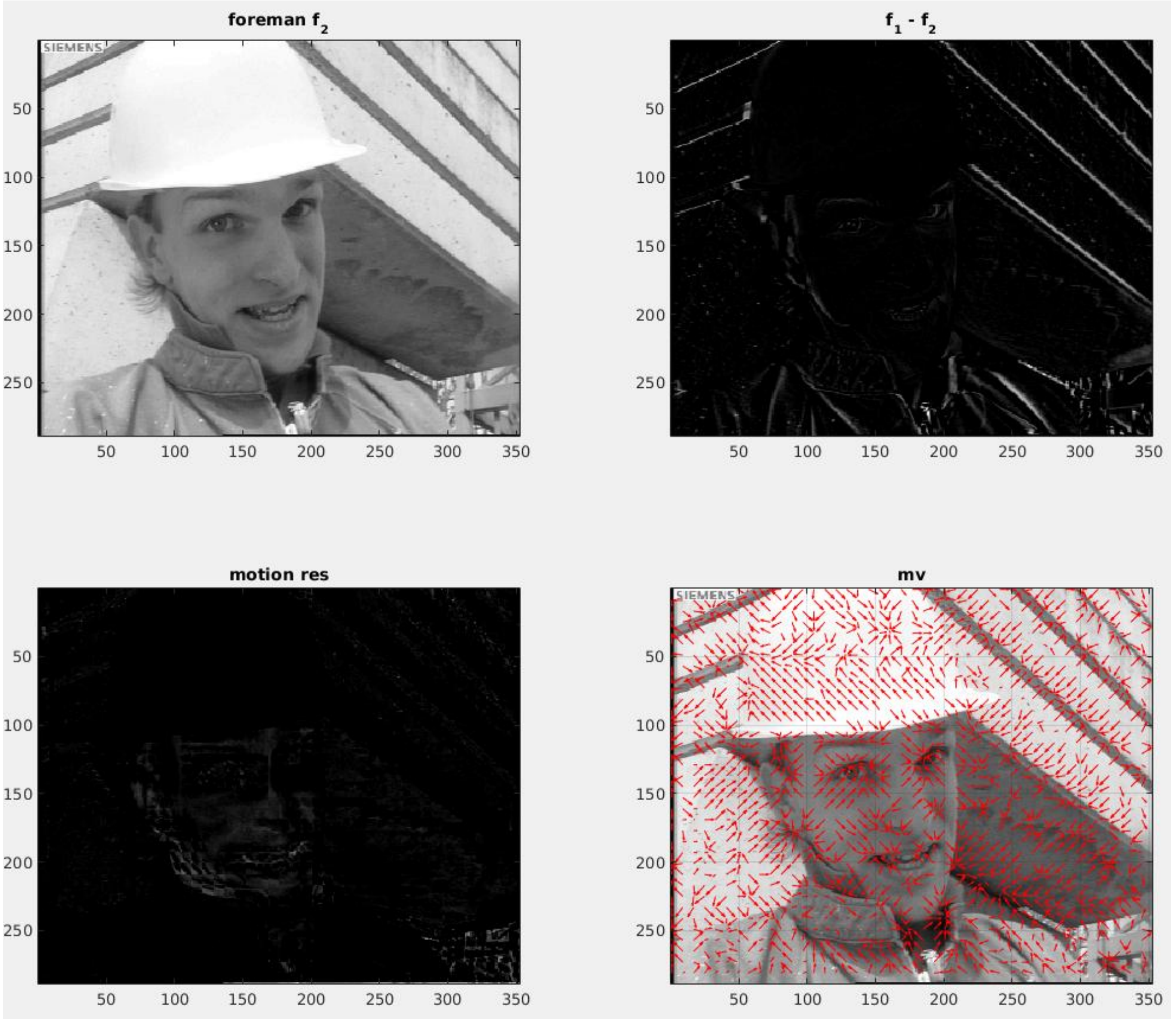
If best match does not lie in the center of large diamond, center large diamond pattern at new best match

Renxiang Li, Bing Zeng, Ming L. Liou:
A new three-step search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Techn. 4(4): 438-442 (1994)

Inter-Prediction

- The purpose: reduce the pixel entropy / variance
- The differential entropy of a pixel is bounded by its Gaussian entropy of the same variance



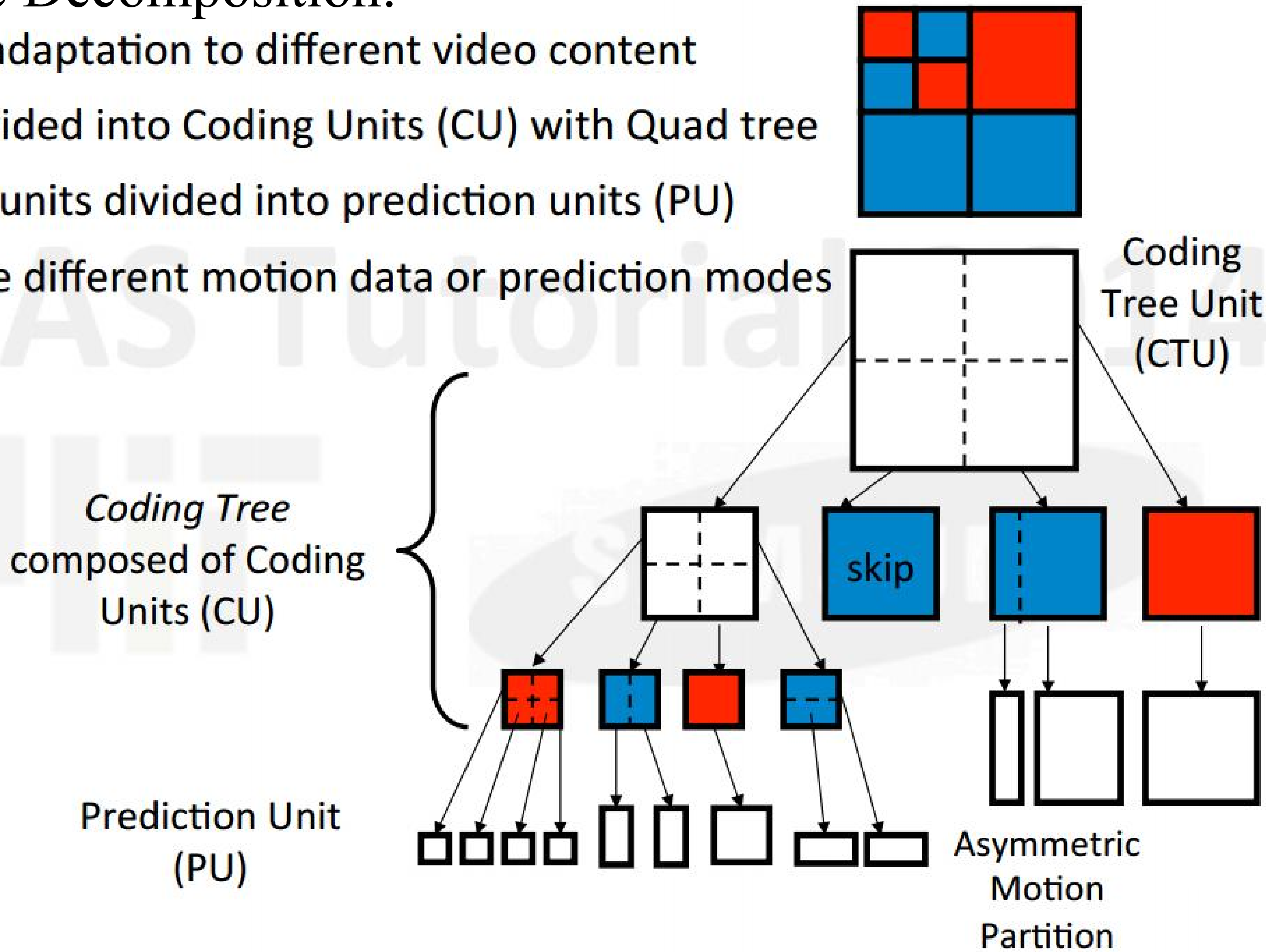


HEVC Coding Structure

Slide Credit: Vivienne Sze & Madhukar Budagavi, ISCAS 2014 Tutorial

❑ Quad Tree Decomposition:

- Better adaptation to different video content
- CTU divided into Coding Units (CU) with Quad tree
- Coding units divided into prediction units (PU)
- PU have different motion data or prediction modes

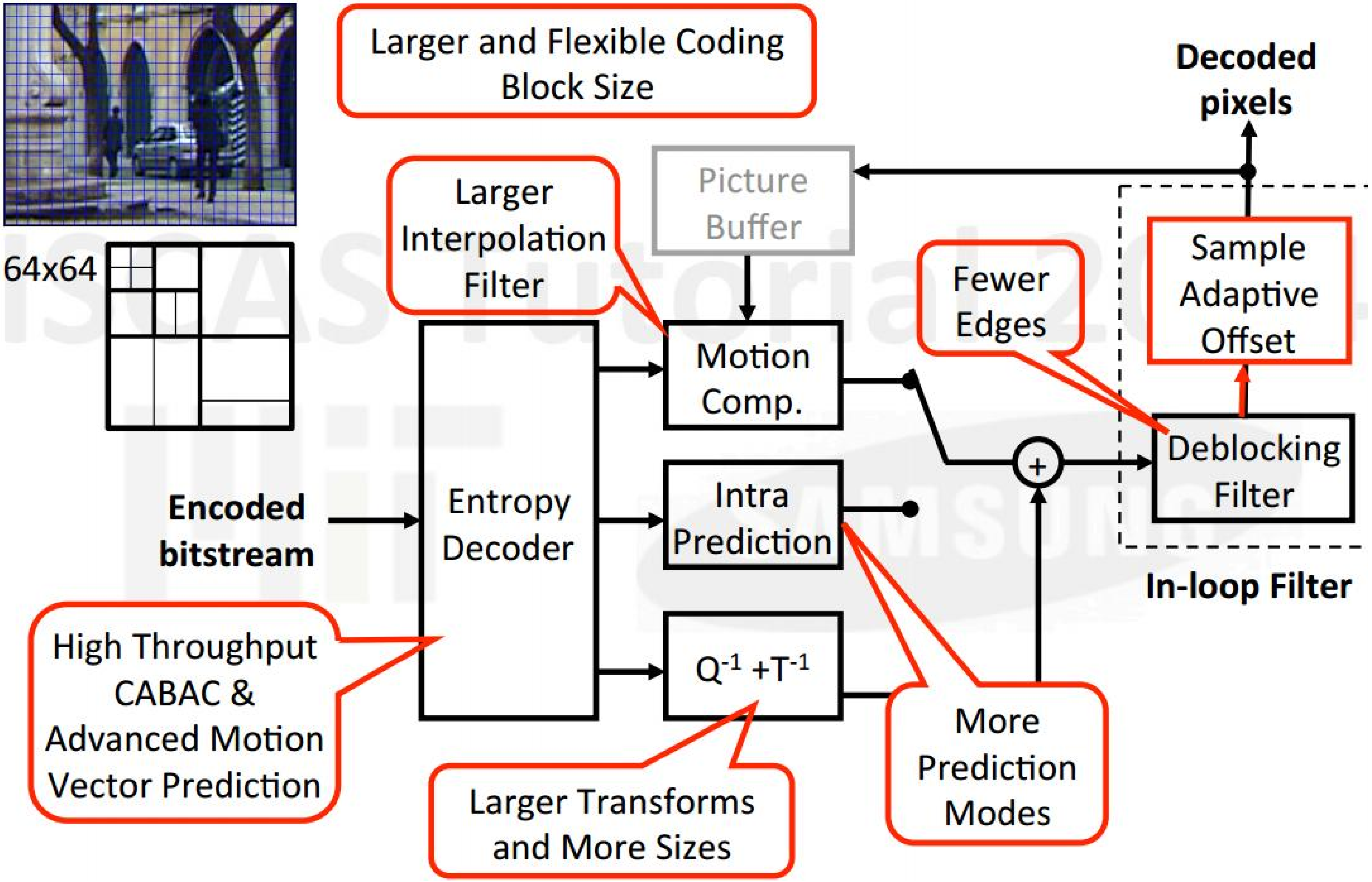


Ref:

G. Schuster, PhD Thesis, 1996: Optimal Allocation of Bits Among Motion, Segmentation and Residual

HEVC Coding Tools

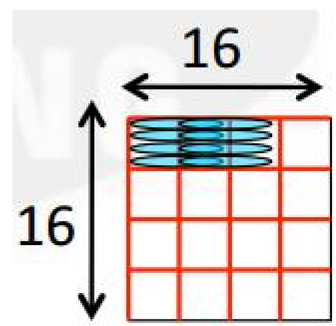
- HEVC (H.265) vs AVC (H.264)



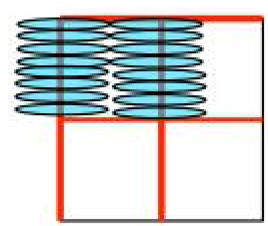
Deblocking Filter

- Reduce blocking artifact in the reconstructed frames
- Can improve both subjective and objective quality
- Filter in H.261:
  - [1/4, 1/2, 1/4]: Applied to non-block-boundary pixels in each block.
  - A low-pass smoothing filter.
- In H.264 (and H.263v2), this is used in the prediction loop to improve motion estimation accuracy. Decoder needs to do the same. Also called loop filter.

Before…. and After



H.264:
4x4 block level



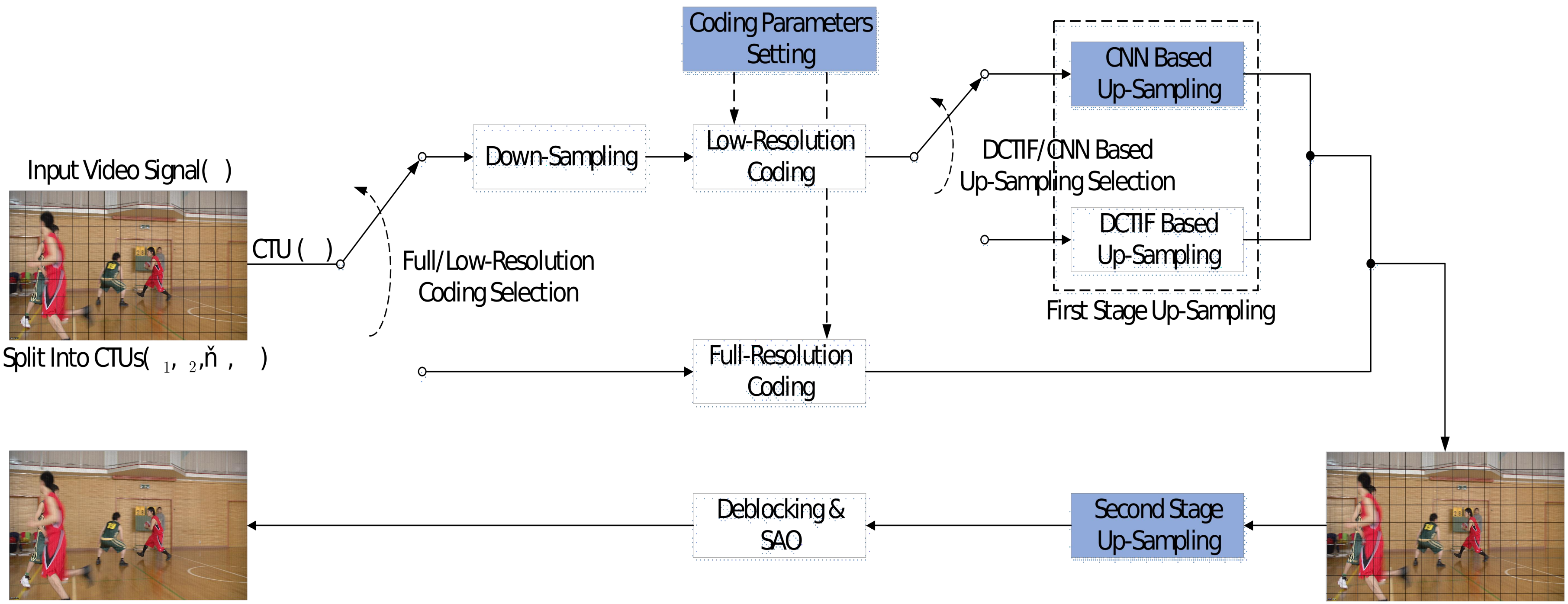
H.265:
8x4 block level

Outline

- Current SOTA Compression Framework Intro
- Deep Learning (DL) as a compression tool in standard based codec
  - DL Super Resolution in intra coding
  - Residual + Reconstruction Deep Learning for deblocking
  - Deep Learning Interpolation in Motion Compensation
  - Deep Learning Chroma Prediction
- End to End Learning based compression
- Summary & Discussions

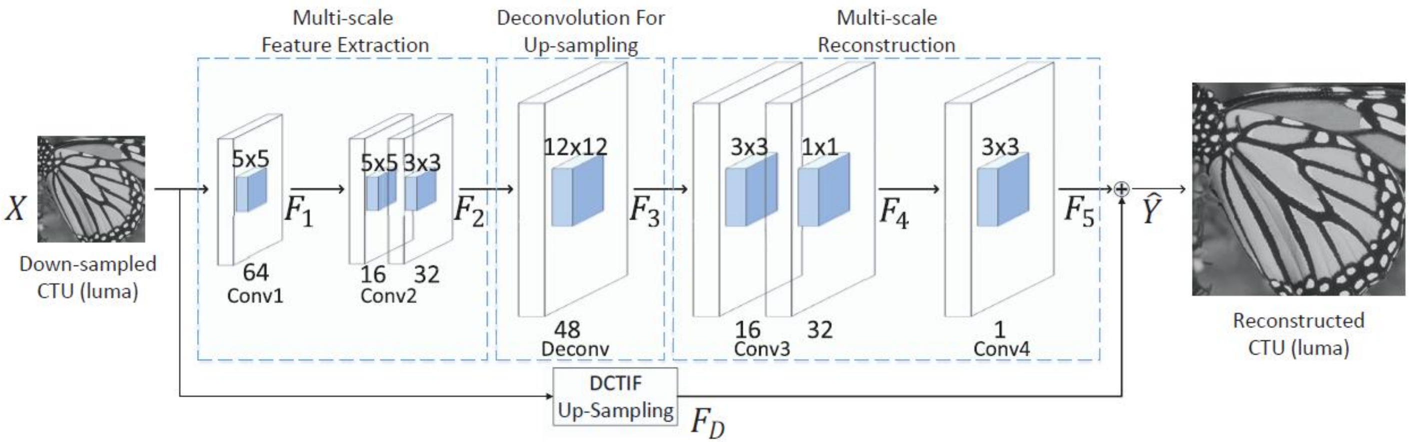
The DLSR encoding framework

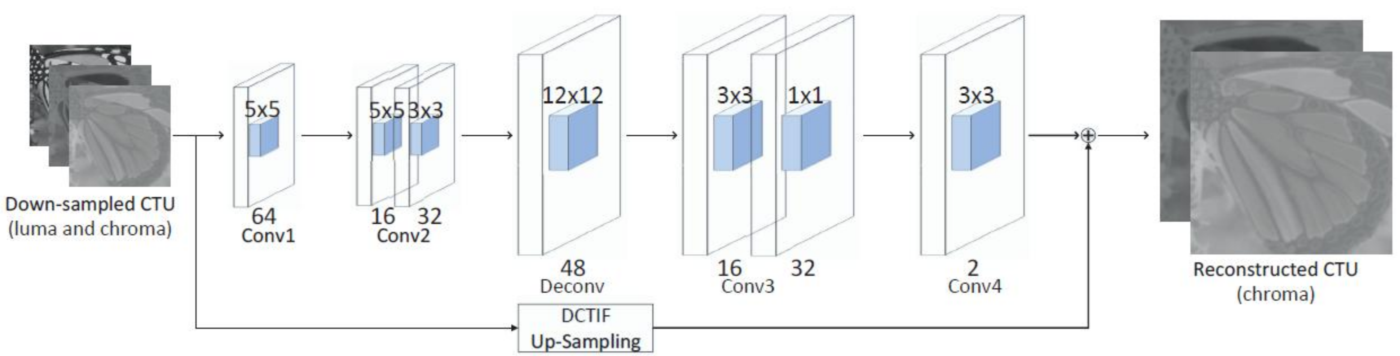
- The framework of the proposed scheme
  - The CNN based Up-Sampling can provide a more accurate pixel or block prediction for pixels



The DCITF Residual Learning SR network

- ❑ We designed a five-layer CNN for the up-sampling of luma and chroma





Experiments

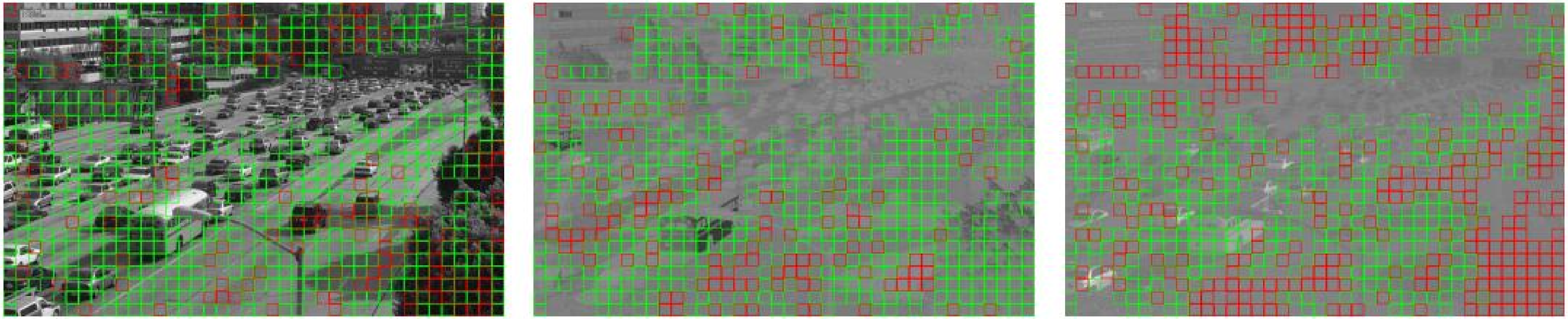
- Performance of the proposed algorithm
  - For UHD test images, we can save up to 9.0% bitrates in average under the same PSNR

Class	Sequence	BD-Rate (Anchored on HEVC)				BD-Rate (Anchored on HEVC+DCTIF)			
		Y	U	V	Y SSIM	Y	U	V	Y SSIM
Class A	Traffic	−10.1%	−3.5%	6.0%	−12.9%	−8.0%	−13.2%	−2.6%	−7.9%
	PeopleOnStreet	−9.7%	−14.8%	−14.5%	−12.9%	−8.5%	−20.4%	−18.5%	−9.7%
	Nebuta	−2.0%	−22.0%	3.1%	−4.4%	−1.7%	−22.5%	1.6%	−3.6%
	SteamLocomotive	−1.7%	−27.7%	−25.4%	−6.1%	−1.2%	−34.2%	-25.6%	−2.8%
Class B	Kimono	−7.7%	−5.5%	18.8%	−9.6%	−3.4%	−25.9%	−4.3%	−3.4%
	ParkScene	−7.1%	−14.4%	−2.3%	−11.3%	−5.0%	−25.2%	−14.6%	−6.6%
	Cactus	−6.6%	−2.5%	8.3%	−10.0%	−5.0%	−6.5%	0.9%	−6.7%
	BQTerrace	−3.7%	−7.6%	−9.1%	−9.6%	−3.1%	−8.2%	−7.1%	−6.5%
	BasketballDrive	−6.1%	−1.2%	3.2%	−10.8%	−3.4%	−5.8%	−2.5%	−3.8%
Class C	BasketballDrill	−4.9%	4.5%	8.1%	−7.9%	−4.0%	4.9%	2.1%	−6.6%
	BQMall	−2.9%	−7.2%	−7.2%	−6.2%	−2.3%	−10.6%	−9.1%	−5.3%
	PartyScene	−1.0%	−5.1%	−1.6%	−4.0%	−1.0%	−5.5%	−3.2%	−3.6%
	RaceHorsesC	−6.7%	4.6%	7.5%	−10.7%	−6.0%	1.9%	3.9%	−8.6%
Class D	BasketballPass	−2.0%	−3.7%	9.2%	−4.3%	−2.3%	−7.5%	12.3%	−4.4%
	BQSquare	−0.9%	−0.6%	−21.1%	−1.4%	−0.5%	1.7%	-16.7%	−1.2%
	BlowingBubbles	−3.2%	3.1%	−8.0%	−5.3%	−1.7%	0.5%	-9.6%	−3.8%
	RaceHorses	−9.9%	7.5%	6.4%	−12.6%	−9.6%	5.0%	6.6%	−11.1%
Class E	FourPeople	−7.2%	−10.5%	−11.0%	−11.0%	−7.2%	−14.7%	−14.5%	−9.5%
	Johnny	−9.0%	−3.2%	−3.2%	−11.1%	−7.1%	−6.0%	−8.3%	−5.6%
	KristenAndSara	−6.8%	−11.2%	−11.1%	−13.0%	−5.3%	−8.4%	−10.6%	−8.2%
Class UHD	Fountains	−4.0%	−12.9%	−11.2%	−7.4%	−2.0%	−16.1%	−9.2%	−2.0%
	Runners	−11.2%	22.8%	−0.1%	−12.4%	−7.0%	0.9%	−13.7%	−6.0%
	Rushhour	−8.5%	4.4%	1.8%	−10.3%	−3.2%	−9.2%	−9.5%	−3.0%
	TrafficFlow	−12.7%	−11.7%	−5.8%	−12.7%	−6.9%	−17.3%	−11.9%	−5.6%
	CampfireParty	−8.4%	−10.8%	−0.8%	−9.5%	−6.5%	−10.8%	−5.0%	−6.4%
Average of Classes A–E		−5.5%	−6.0%	−2.2%	−8.8%	−4.3%	−10.0%	−6.0%	−5.9%
Average of Class UHD		−9.0%	−1.6%	−3.2%	−10.5%	−5.1%	−10.5%	−9.9%	−4.6%

Experiments

- Ratio of the DL methods activated in the operation
  - $P_{hitting}$: Ratio of DLSR in RDO modes
  - P_{luma}, P_{Cb}, P_{Cr}: Ratio of CNN

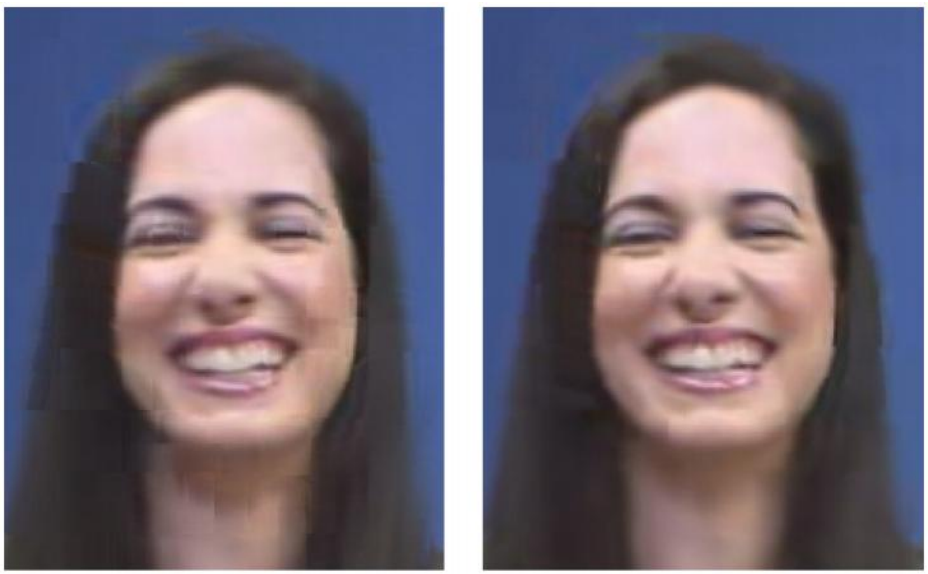
Class	$P_{Hitting}$	P_{Luma}	P_{Cb}	P_{Cr}
Class A	72.2%	70.3%	71.2%	55.0%
Class B	68.4%	75.0%	65.1%	49.4%
Class C	48.1%	92.0%	68.5%	73.5%
Class D	42.4%	81.9%	51.6%	70.7%
Class E	68.7%	72.8%	54.4%	58.5%
Class UHD	85.2%	68.4%	54.2%	64.1%



HEVC in-loop filter

- The loop filters in HEVC

De-blocking filter



Sample adaptive offset

- Advantages: low complexity
- Disadvantages: fixed, limited performance improvement

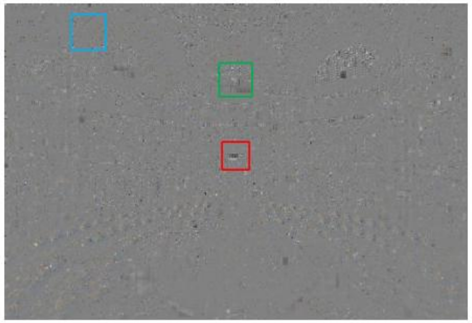
Motivation - Guided Filtering

- The residual frame can be used as the guidance for the in-loop filter of the reconstructed frame
  - Larger residuals indicate larger reconstruction errors

(a) Origin

(b) Reconstruction

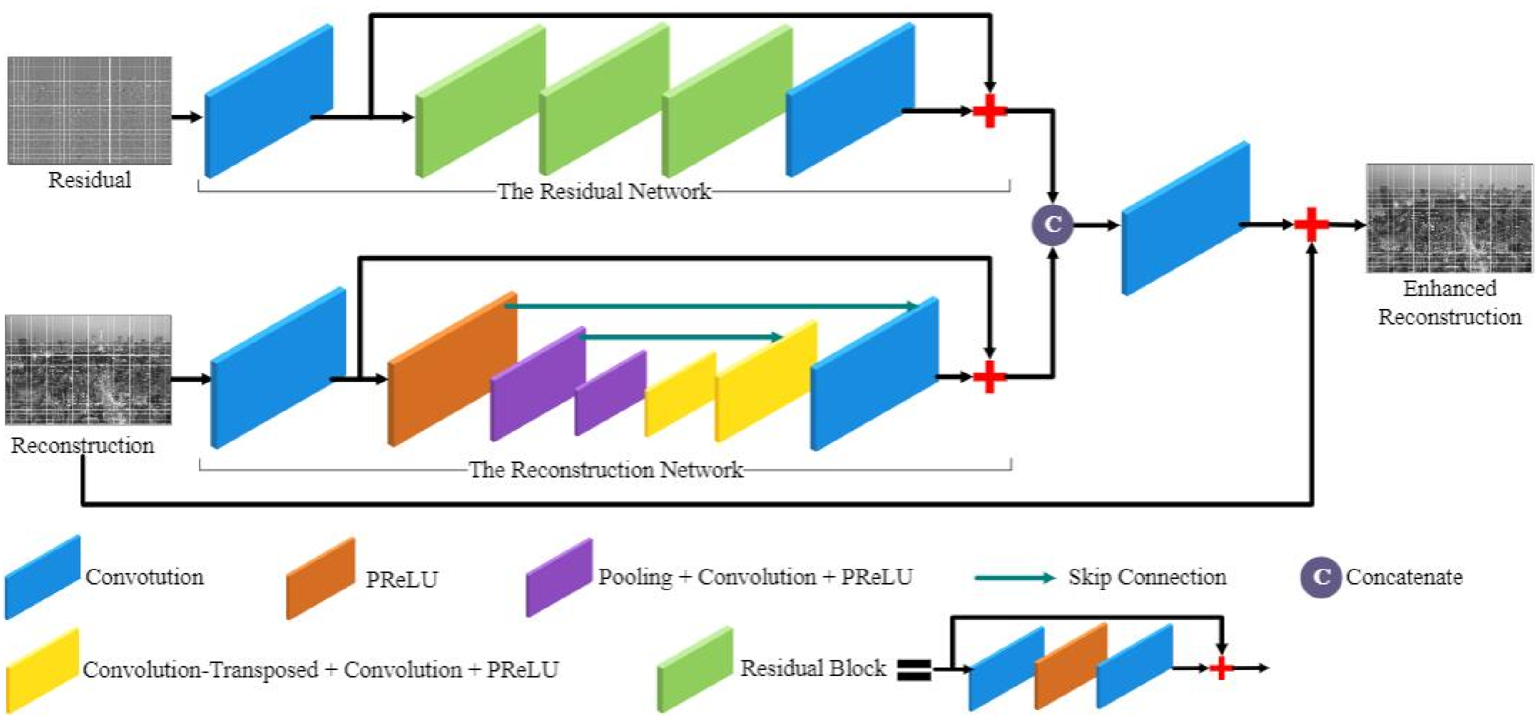
(c) Prediction



(d) Residual

Coding-prior-based in-loop filter

- The residual frame is used as the additional input
- Specific networks for reconstruction and residual
  - Residual Network: residual blocks
  - Reconstruction Network: down-sampling and up-sampling



Experimental results

❑ Comparison with VRCNN

Intra: 2.1% improvement

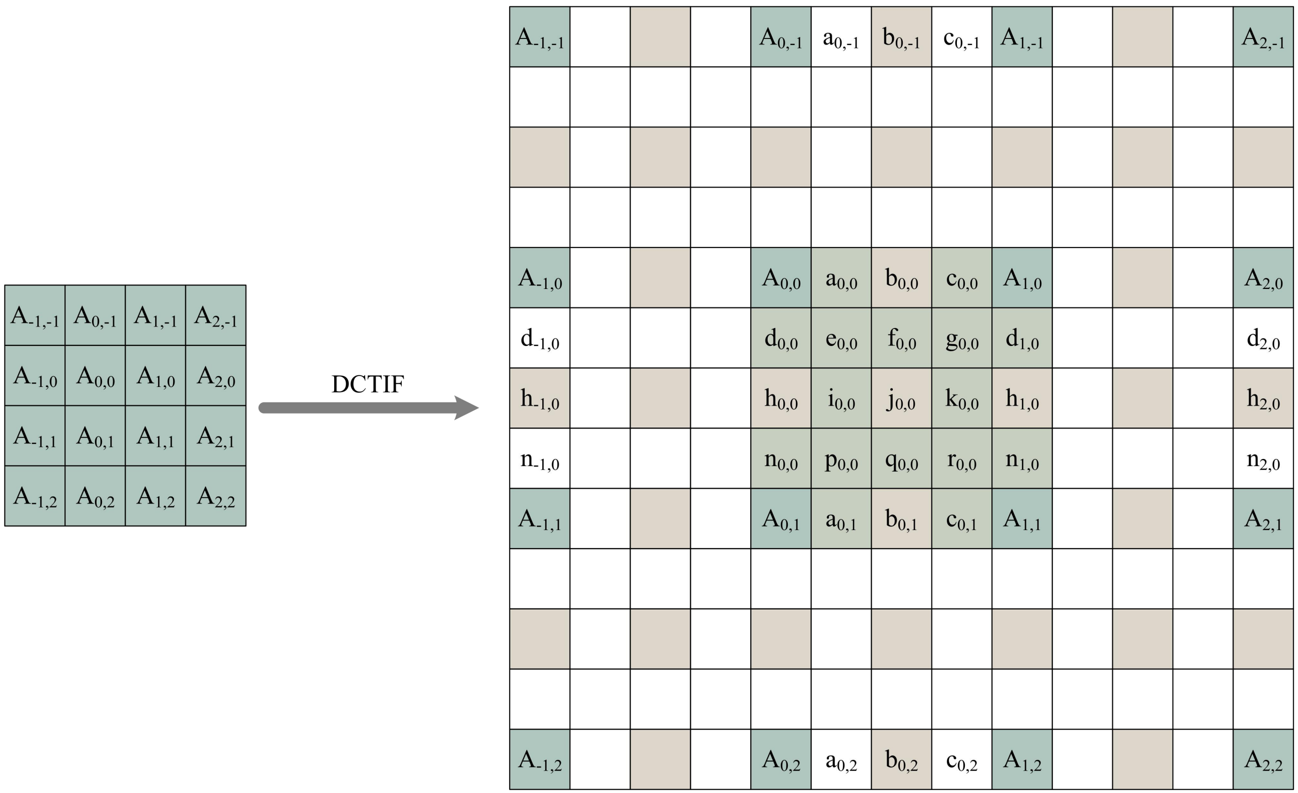
Class	Sequence	VRCNN vs. HEVC	RRCNN vs. HEVC
A	Traffic	−8.1%	**-10.2%**
	PeopleOnStreet	−7.7%	**-9.4%**
B	Kimono	−5.9%	**-8.6%**
	ParkScene	−6.2%	**-8.1%**
	Cactus	−2.7%	**-5.8%**
	BasketballDrive	−5.2%	**-7.7%**
	BQTerrace	−2.9%	**-4.2%**
C	BasketballDrill	−10.6%	**-13.8%**
	BQMall	−7.3%	**-9.3%**
	PartyScene	−4.6%	**-5.6%**
	RaceHorses	−5.8%	**-7.1%**
D	BasketballPass	−7.6%	**-9.5%**
	BQSquare	−5.3%	**-6.3%**
	BlowingBubbles	−5.5%	**-6.7%**
	RaceHorses	−8.9%	**-10.2%**
E	FourPeople	−10.0%	**-12.8%**
	Johnny	−9.1%	**-12.5%**
	KristenAndSara	−9.4%	**-11.8%**
	Class A	−7.9%	**-9.8%**
	Class B	−4.6%	**-6.9%**
	Class C	−7.1%	**-8.9%**
	Class D	−6.8%	**-8.2%**
	Class E	−9.5%	**-12.4%**
Avg.	All	−6.8%	**-8.9%**

Inter: 0.7% improvement

Class	Sequence	VRCNN vs. HEVC	RRCNN vs. HEVC
A	Traffic	−5.0%	**-6.0%**
	PeopleOnStreet	−1.4%	**-1.6%**
B	Kimono	−1.9%	**-2.6%**
	ParkScene	−2.7%	**-3.4%**
	Cactus	−3.2%	**-3.9%**
	BasketballDrive	−1.4%	**-1.9%**
	BQTerrace	−5.2%	**-5.8%**
C	BasketballDrill	−3.1%	**-4.3%**
	BQMall	−2.0%	**-2.5%**
	PartyScene	−0.5%	**-1.0%**
	RaceHorses	−1.3%	**-1.4%**
D	BasketballPass	−0.7%	**-0.9%**
	BQSquare	−1.4%	**-2.1%**
	BlowingBubbles	−1.8%	**-2.4%**
	RaceHorses	−1.5%	**-1.6%**
E	FourPeople	−8.2%	**-9.5%**
	Johnny	−7.6%	**-10.2%**
	KristenAndSara	−6.9%	**-7.6%**
	Class A	−3.2%	**-3.8%**
	Class B	−2.9%	**-3.5%**
	Class C	−1.7%	**-2.3%**
	Class D	−1.4%	**-1.7%**
	Class E	−7.6%	**-9.1%**
Avg.	All	−3.1%	**-3.8%**

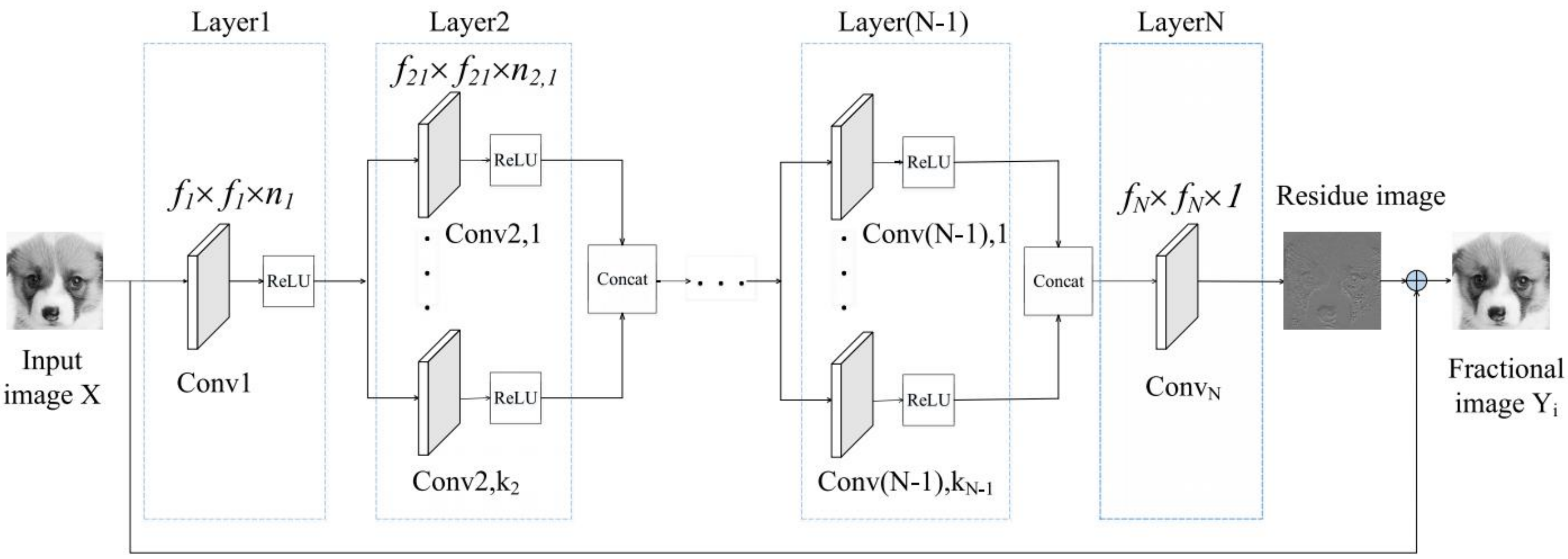
HEVC interpolation filter

- DCTIF: fixed and simple interpolation filter
  - Luma: 8/7 taps interpolation filter
  - Chroma: 4 taps interpolation filter



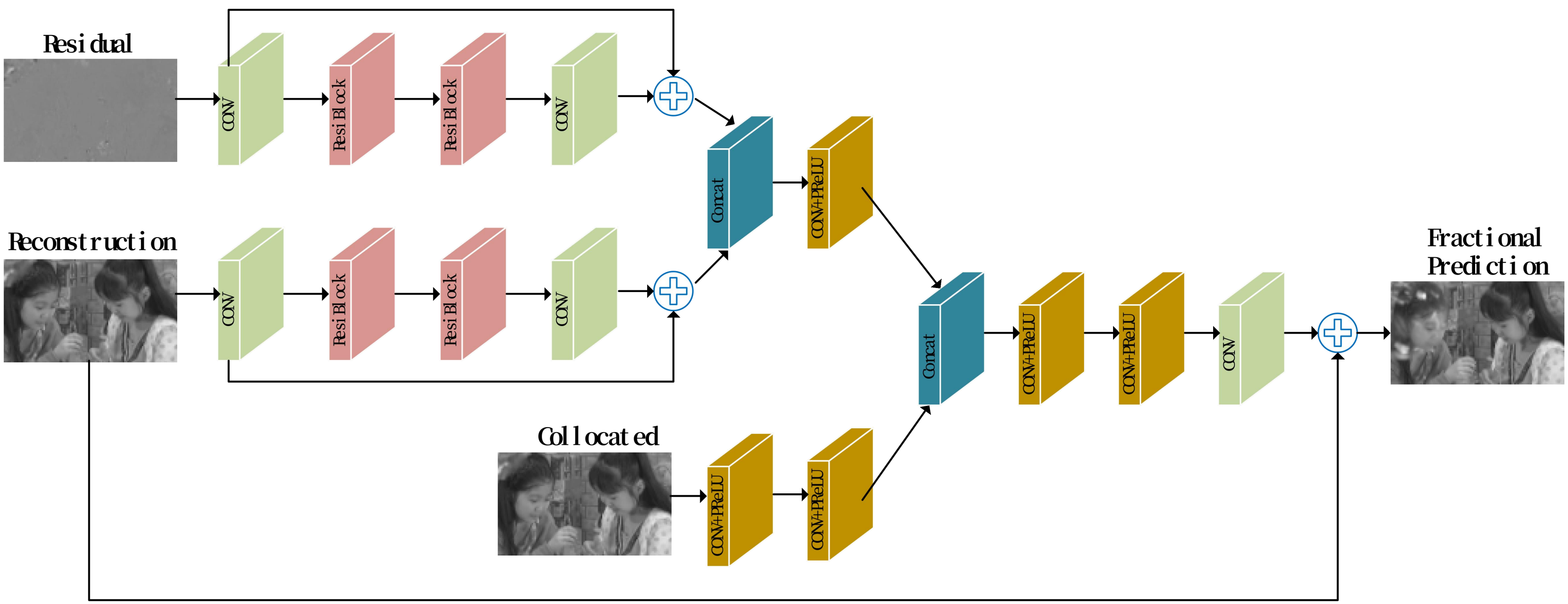
Deep-learning-based interpolation filter

- CNN-based fractional pixel motion compensation
  - Input: reconstructed block
  - Label: original block
  - No coding-prior information during compression is used



Rich Coding Prior Deep Learning Interpolation

- The residual block and the co-located high quality blocks are used as the additional inputs of the CNN
- We design specific network structures for the residual, reconstruction and collocated blocks



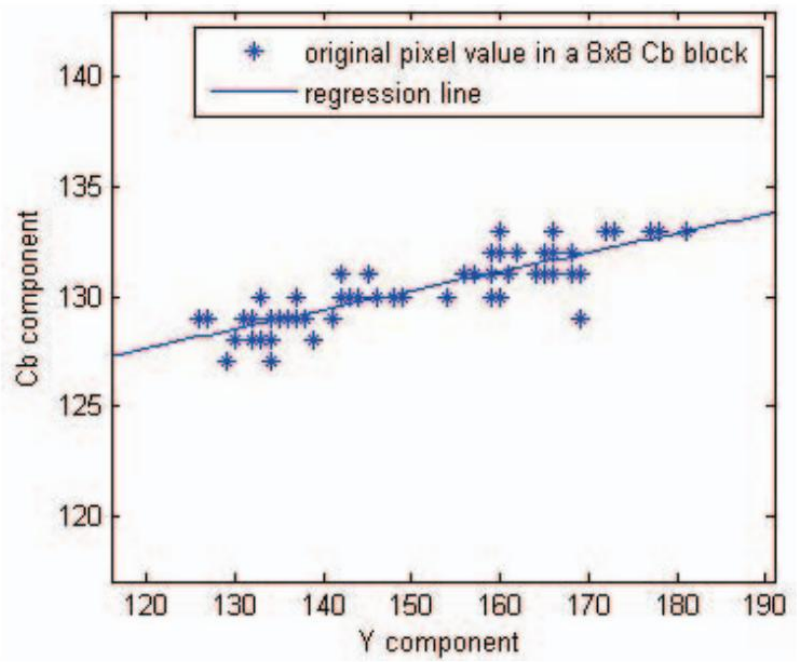
Experimental results

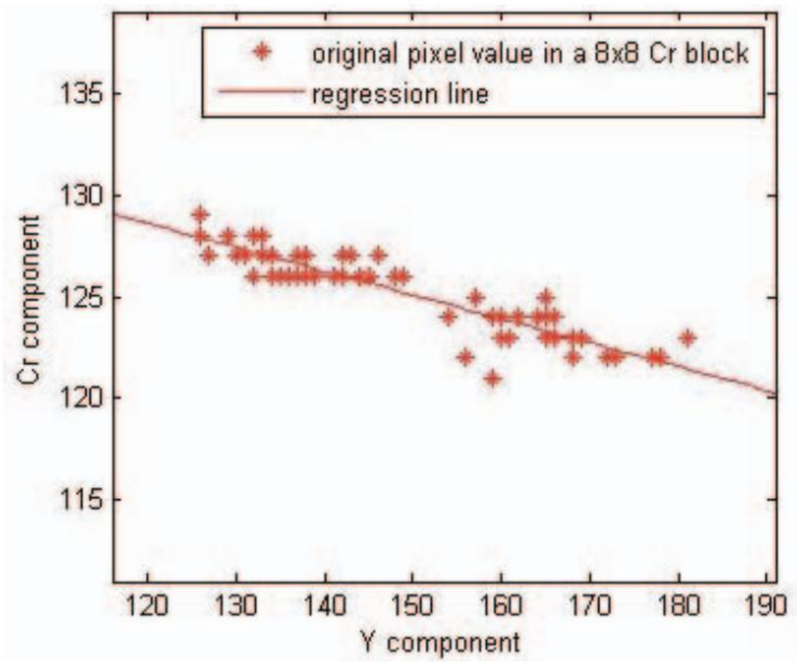
- The coding prior can bring 5.3% improvement over HEVC

Class	Sequence	BD-rate of LDP(%)		
		FRCNN[1]	GVCNN[2]	Proposed
Class B	Kimono	-4.3	-4.1	-6.7
	ParkScene	-1.9	-5.2	-3.2
	Cactus	-3.8	-3.3	-5.9
	BasketballDrive	-5.0	-1.3	-6.2
	BQTerrace	-6.5	-2.5	-11.4
Class C	BasketballDrill	-4.0	-2.2	-5.1
	BQMall	-4.8	-2.9	-5.5
	PartyScene	-3.2	-1.6	-3.6
	RaceHorses	-3.0	-2.0	-4.0
Class D	BasketballPass	-3.3	-3.3	-4.8
	BQSquare	-4.2	-2.1	-5.6
	BlowingBubbles	-4.7	-0.6	-4.4
	RaceHorses	-1.9	-2.7	-3.8
ClassE	FourPeople	-5.7	-1.6	-8.3
	Johnny	-6.2	-2.9	-9.0
	KristenAndSara	-6.3	-2.2	-8.4
ClassF	BasketballDrillText	-4.1	-1.8	-4.8
	ChinaSpeed	-2.0	-1.4	-1.7
	SlideEditing	-0.7	0.0	-0.3
	SlideShow	-2.3	-0.5	-2.3
Average	ClassB	-4.3	-3.3	-6.7
	ClassC	-3.8	-2.2	-4.5
	ClassD	-3.5	-2.2	-4.6
	ClassE	-6.1	-2.2	-8.5
	ClassF	-2.3	-0.9	-2.3
Overall	All Sequences	-3.9	-2.2	-5.3

Linear prediction from Luma to Chroma

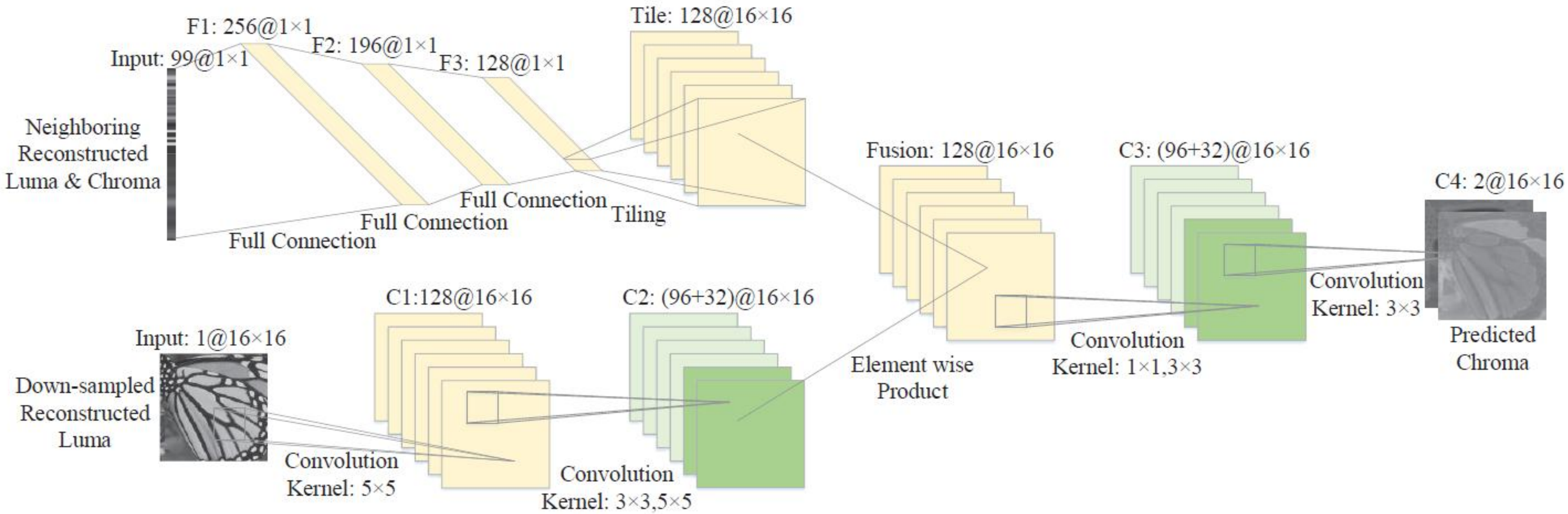
- Linear relationship between Y, Cb, and Cr
  - Luma down-sampling $y'_R(i,j) = [y_R(2i,2j) + y_R(2i,2j+1)]/2$
  - Linear relationship $c_p(i,j) = \alpha * y'_R(i,j) + \beta; i,j = 0, ..., N-1$





Chroma Prediction in Coding

- The coding priors neighboring luma and chroma pixels are used as the additional input
  - Neighboring pixels: full connected layers
  - Current luma block: convolutional layers



Experimental results

- 0.2%, 3.1%, and 2.0% performance improvements on Y, U, and V components

Class	Sequence	Y	U	V
Class A	Traffic	–0.0%	–2.1%	–0.7%
	PeopleOnStreet	–0.2%	–2.4%	–2.5%
	Nebuta	–0.6%	–9.2%	–0.8%
	SteamLocomotive	–0.0%	–9.1%	1.5%
Class B	Kimono	–1.2%	–5.4%	0.1%
	ParkScene	–0.7%	–8.9%	–0.7%
	Cactus	–0.0%	–4.0%	–3.8%
	BQTerrace	–0.2%	0.4%	–1.9%
	BasketballDrive	–0.2%	–3.6%	–4.2%
Class C	BasketballDrill	–0.3%	–1.6%	0.7%
	BQMall	–0.1%	–5.7%	–3.4%
	PartyScene	–0.2%	–4.6%	–0.9%
	RaceHorsesC	–0.1%	0.6%	–0.3%
Class D	BasketballPass	–0.9%	–1.3%	–4.2%
	BQSquare	0.4%	0.6%	–0.3%
	BlowingBubbles	0.4%	–3.2%	–7.8%
	RaceHorses	0.4%	–2.0%	–1.8%
Class E	FourPeople	–0.1%	–1.8%	–0.2%
	Johnny	0.6%	0.4%	–3.8%
	KristenAndSara	–0.4%	1.2%	–5.2%
Average		–0.2%	–3.1%	–2.0%

References

- Some of the recent and on-going work

  - Y. Li, L. Li, D. Li, H. Li, Z. Li, and F. Wu, “Learning a Convolutional Neural Network for Image Compact Resolution”, accepted, IEEE Trans on Image Processing , 2018.
  - Y. Li, L. Li, Z. Li, J. Yang, N. Xu, D. Liu, H. Li, "A Hybrid Neural Network for Chroma Intra Prediction" IEEE Int'l Conf on Image Processing (ICIP), Athens, Greece, 2018.
  - H. Zhang, L. Li, L. Song, X.-K. Yang, Z. Li "Advanced CNN Based Motion Compensation Fractional Interpolation", IEEE Int'l Conf on Image Processing (ICIP), 2019.
  - Z. Zhang, L. Li, Z. Li, and H. Li, "Mobile Visual Search Compression with Grassmann Manifold Embedding", IEEE Trans on Circuits & Sys for Video Tech , accepted.

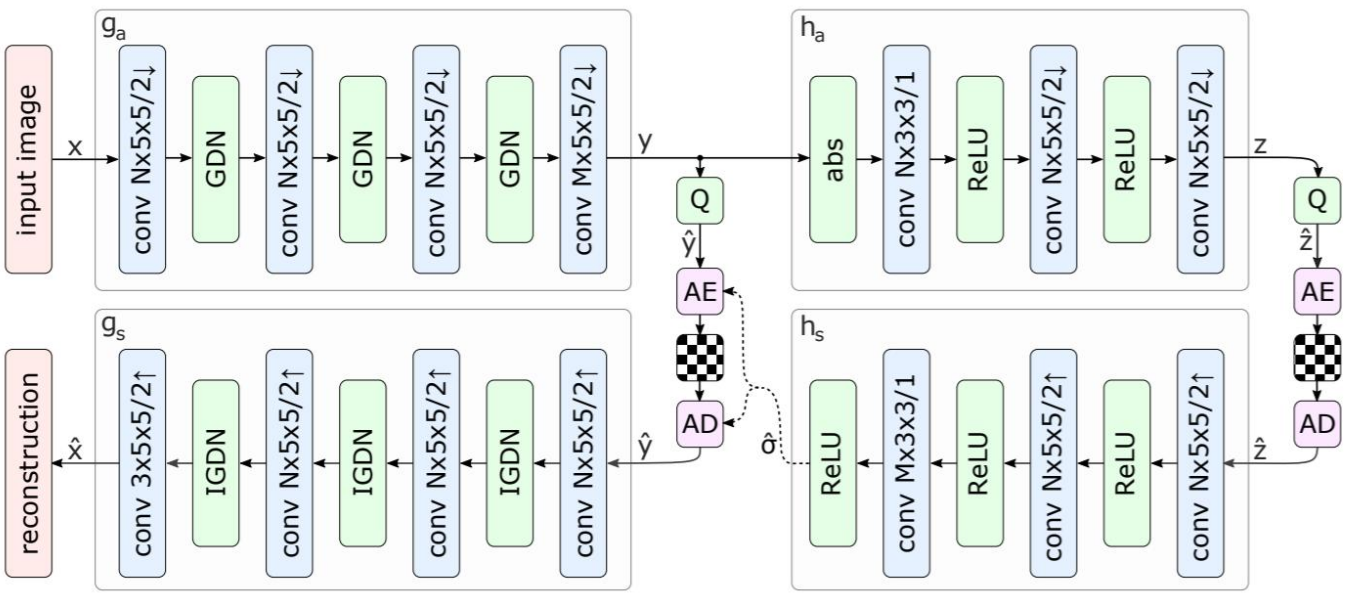
Outline

- Current SOTA Compression Framework Intro
- Deep Learning as a compression tool in standard based codec
- End to End Learning based compression
- Summary & Discussions

End-End Deep Learning Based

❑ Motivation

- Deep learning for compression has achieved remarkable progress and attracted quite some attention.
- Google's Variational Autoencoder:
  - differentiable quantization loss via AWGN type noise
  - context model for the Arithmetic Coding



Current SOTA Models

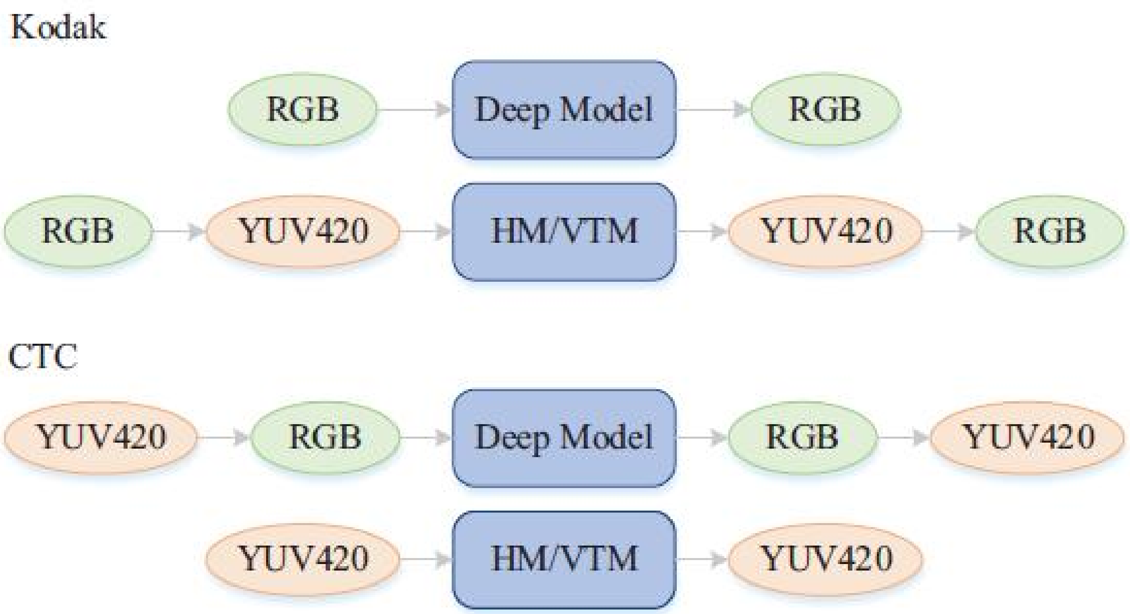
- Learning-based Methods
  - G. Toderici *et al.*, 2017
  - J. Balle, 2018
  - J. Balle, *et al.*, 2018
  - D. Minnen *et al.*, 2018
  - F. Mentzer *et al.*, 2018
  - J. Lee *et al.*, 2019

Method	Implementations
G. Toderici *et al.*, 2017 [10]	rnn-compression [18]
J. Ballé, 2018 [4] J. Ballé *et al.*, 2018 [6] D. Minnen *et al.*, 2018 [7]	Tensorflow Data Compression [19]
F. Mentzer *et al.*, 2018 [9]	imgcomp-cvpr [20]
J. Lee *et al.*, 2019 [8]	CA_Entropy_Model [8]

- Standards based Codecs
  - HEVC
  - VVC

Datasets & Common Test Condition (CTC)

- Kodak[1]
- VVC CTC sequences
- For fair comparison, evaluate in the same color space



The PCS 2013
VP9 Scandal !

Model Configurations

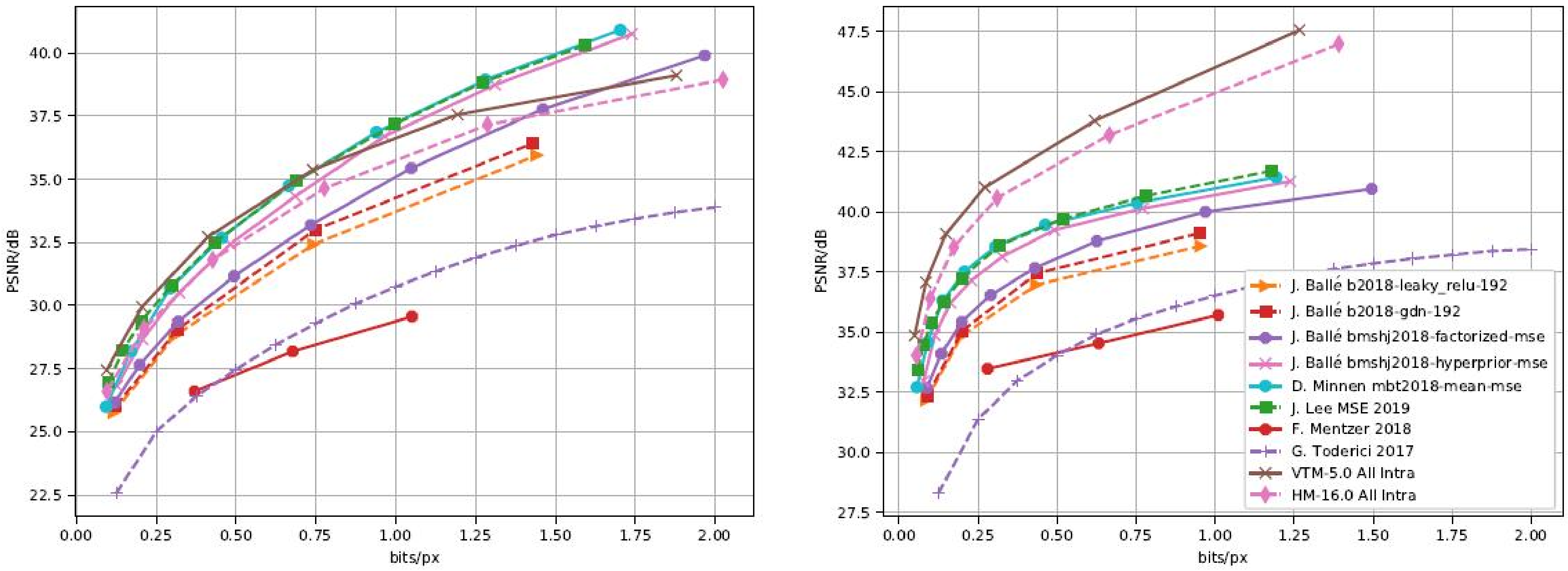
- Learning-based methods
  - Using Tensorflow repository from their papers

- HEVC and VVC
  - HM 16.0
  - VTM 5.0
  - QP = [17, 42, 22, 27, 32, 37, 42]
  - Bit depth = 8

Accessible at: http://www.cipr.rpi.edu/resource/stills/kodak.html

PSNR Results

- On Kodak dataset, VTM performs best in <0.75 bpp area, while is surpassed by learning-based method when bpp is larger than 0.75.
- On CTC sequences, HM/VTM are significantly better.

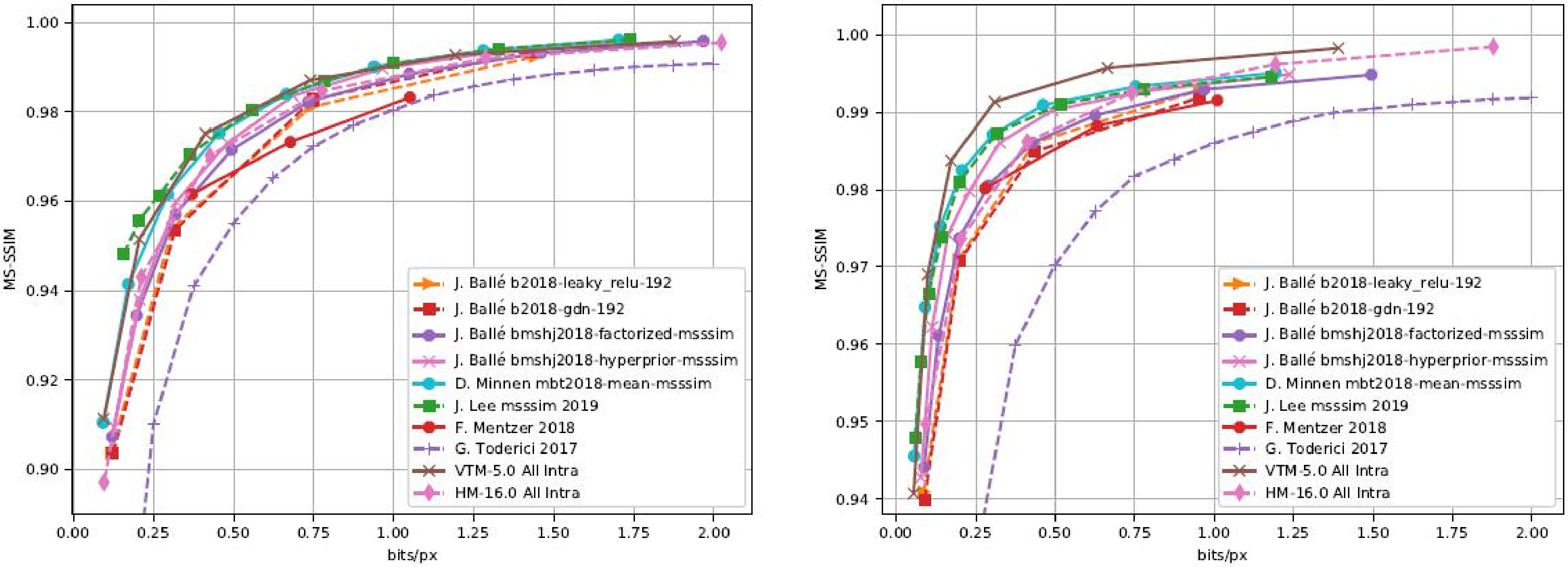


(a) Kodak dataset. (b) VVC CTC test sequences.

Fig. 2: PSNR results using HM/VTM comparing with LBC methods on Kodak and CTC test sequences.

MS-SSIM Results

- On Kodak dataset, VTM and LBC are comparable.
- On CTC sequences, HM/VTM are significantly better.



(a) Kodak dataset. (b) VVC CTC test sequences.

Fig. 3: MS-SSIM results using HM/VTM comparing with LBC methods on Kodak and CTC test sequences.

Time Complexity

Table 2: Decoding time in seconds on VVC CTC test sequences.

Class	Sequence Name	J. Ballé [4]		J. Ballé [6]		D. Minnen [7]	J. Lee [8]	F. Mentzer [9]	G. Toderici [10]	VTM-5.0	HM-16.0
		LReLU	GDN	factorized	hyperprior						
A1	Tango2	6.655	6.346	9.596	9.368	14.289	873.163	36.908	14.493	0.645	0.827
	FoodMarket4	6.127	5.706	8.847	8.764	13.273	786.189	27.384	9.436	0.554	0.697
	Campfire	5.245	5.462	8.341	8.377	13.509	772.247	33.829	5.292	1.085	0.971
A2	CatRobot	6.106	5.914	9.695	9.032	13.828	790.025	27.630	10.443	0.792	0.894
	DaylightRoad2	6.031	5.719	8.690	8.755	13.378	809.819	27.615	16.271	0.828	0.884
	ParkRunning3	6.063	6.125	9.025	9.034	13.967	830.452	27.319	6.396	1.022	1.152
B	MarketPlace	2.132	2.131	2.919	2.976	4.365	216.086	6.168	1.560	0.227	0.325
	RitualDance	1.992	1.888	2.709	2.748	4.125	206.234	6.190	1.494	0.175	0.232
	Cactus	2.041	2.050	2.771	2.877	4.287	197.209	6.123	4.885	0.353	0.280
	BasketballDrive	2.113	2.012	2.819	2.868	4.269	199.110	6.198	1.726	0.246	0.133
	BQTerrace	1.980	2.008	2.777	3.041	4.260	194.927	6.246	1.481	0.446	0.213
C	RaceHorses	0.914	0.926	1.091	1.144	1.672	46.312	0.803	0.423	0.090	0.062
	BQMall	0.957	0.960	1.129	1.207	1.687	43.170	1.152	0.447	0.090	0.126
	PartyScene	0.912	0.931	1.098	1.161	1.689	41.690	0.806	0.376	0.110	0.198
	BasketballDrill	0.933	0.940	1.099	1.190	1.686	40.976	0.850	0.444	0.100	0.117
D	RaceHorses	0.729	0.755	0.802	0.912	1.196	13.251	0.173	0.249	0.028	0.037
	BQSquare	0.729	0.762	0.805	1.107	1.211	13.268	0.294	0.248	0.036	0.072
	BlowingBubbles	0.731	0.735	0.773	0.841	1.200	12.582	0.167	0.249	0.035	0.042
	BasketballPass	0.751	0.767	0.793	0.874	1.228	12.643	0.185	0.260	0.029	0.048
E	FourPeople	1.316	1.284	1.639	1.769	2.499	102.818	2.796	0.833	0.119	0.141
	Johnny	1.300	1.274	1.627	1.742	2.424	98.666	2.034	0.878	0.090	0.110
	KristenAndSara	1.312	1.309	1.637	1.676	2.446	97.198	2.103	0.852	0.099	0.110
F	ArenaOfValor	2.101	20.534	21.608	26.660	39.328	214.568	8.632	4.544	0.412	0.585
	BasketballDrillText	0.948	1.060	1.116	1.222	1.689	41.446	0.829	0.446	0.081	0.100
	SlideEditing	1.229	1.279	1.606	1.685	2.541	93.765	2.042	0.720	0.126	0.160
	SlideShow	1.230	1.237	1.590	1.616	2.461	94.557	2.046	0.738	0.098	0.219
Total Time		62.577	80.204	106.602	112.646	168.507	6842.371	236.519	85.184	7.956	8.735
Time Complexity*		**7.16**×	**9.18**×	**12.20**×	**12.90**×	**19.29**×	**783.33**×	**27.08**×	**9.75**×	**0.91**×†	**1.00**×

*Compared with HEVC reference software HM-16.0. †SIMD is enabled.

Observations & Discussions

- Learning Based Compression efficiency is still far lagging behind the standard based SOTA codec in image compression, 2-3dB is like 10 years' technology gap
- Complexity of the CNN is prohibitive for current hardware technology for meaningful deployment
- Learning based solution is a good framework for reducing signal differential entropy, indeed standards based solution now involves many modes of opeartions that is equivalent to the many signal paths in the CNN
- Should have a more rigorous grand challenge CTC scheme for learning based compression research.