Deep Neural Network Based Frame Reconstruction For Optimized Video Coding - An AV2 Approach

Dandan Ding
Hangzhou Normal University

ALLIANCE FOR OPEN MEDIA RESEARCH
Symposium 2019

01 Background of our project

AV1 encode preset	Speed vs. VP9 [--cpu-used=0 --auto-alt-ref=6]	BDRATE vs. VP9 [--cpu-used=0 --auto-alt-ref=6]
AV1 [--cpu-used=0]	14.8x	-31.42%
AV1 [--cpu-used=1]	5.3x	-29.24%
AV1 [--cpu-used=2]	3.8x	-26.90%
AV1 [--cpu-used=3]	1.9x	-24.52%
AV1 [--cpu-used=4]	1.6x	-23.10%
AV1 [--cpu-used=5]	1.4x	-21.89%

Mid resolution

Baseline HEVC (HM16.17)	AV1 (libaom bff6ee33)	VVC (VTM 6.0)
Av-PSNR	-20.8%	-25.3%
Glb-PSNR	-21.8%	-25.3%
SSIM	-24.5%	-27.7%

High resolution

Baseline HEVC (HM16.17)	AV1 (libaom bff6ee33)	VVC (VTM 6.0)
Av-PSNR	-24.4%	-28.5%
Glb-PSNR	-26.1%	-28.5%
SSIM	-28.0%	-31.2%

Debargha Mukherjee, Preliminary comparison of AV1 with emergent VVC standard, *ICIP*, 2019.

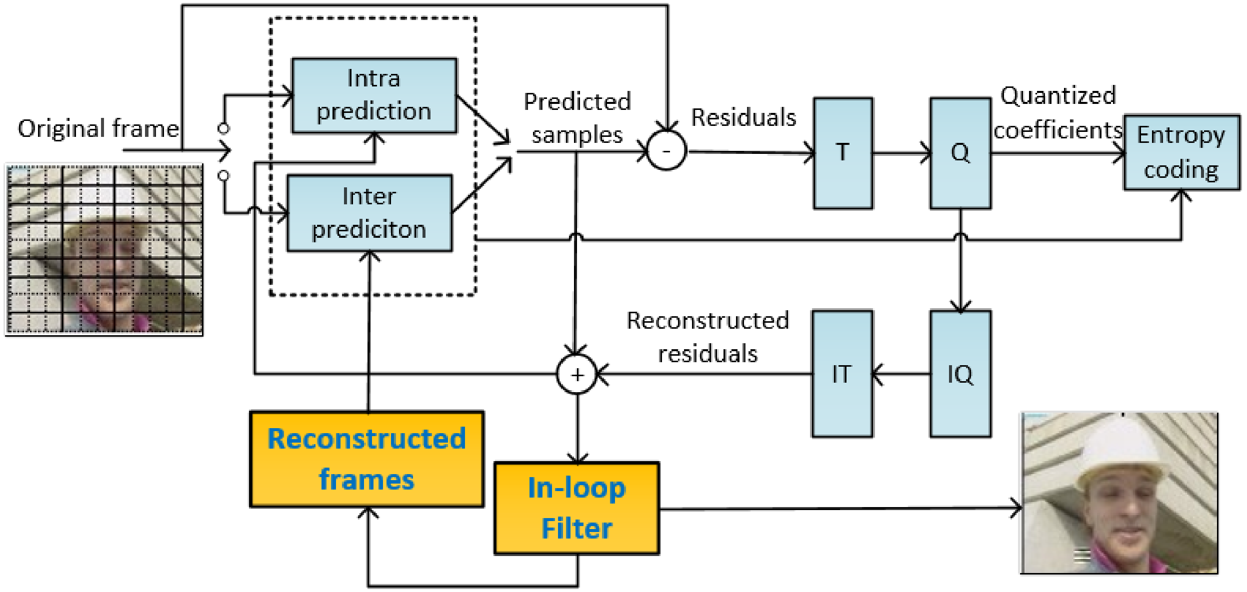
- AV1 is the *most advanced standardized codec* available today.
- Research and development of tools towards a potential successor to AV1, so called AV2, have started.

A viable successor for further BDRATE reduction over AV1.

02 Our Goal

We completely focus on the optimization of reconstruction frames through using the Deep Neural Network (DNN).

In-loop filter



03 Two problems are concerned

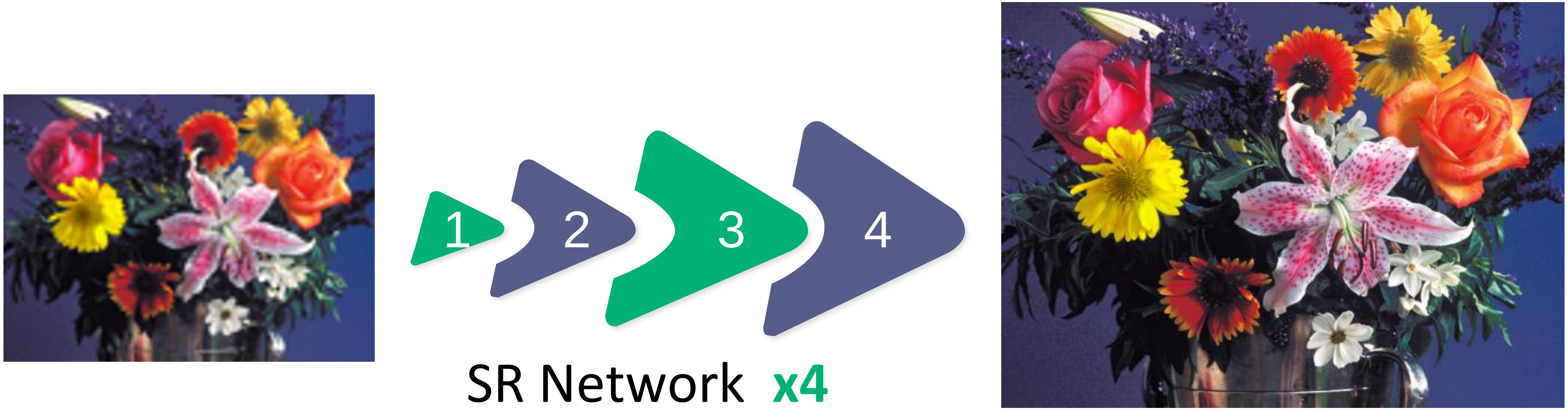
Two aspects are explored, including:

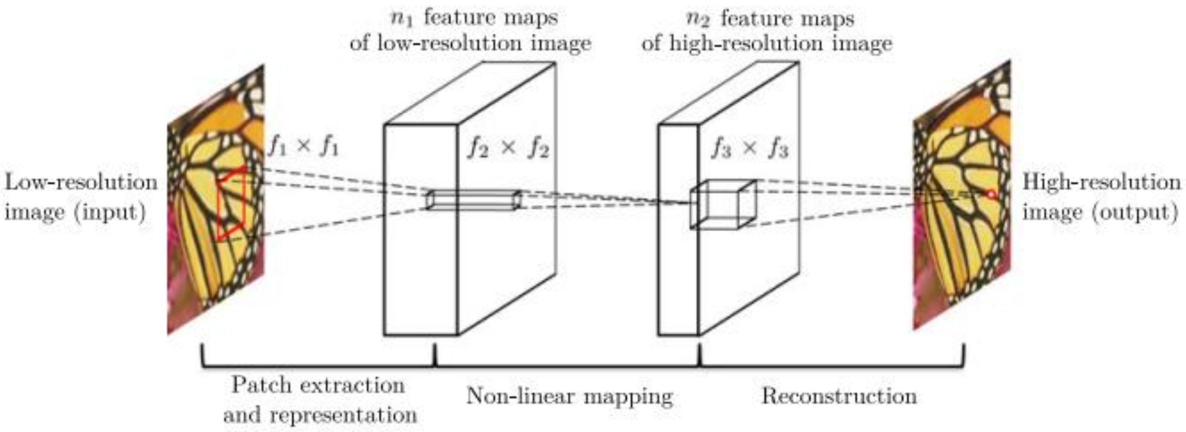
Q1 How to design a CNN-based in-loop filter for AV1?

Q2 How to incorporate the CNN-based filters into AV1 encoder?

Q1 How to design a CNN-based in-loop filter for AV1?

- The problem has similarities with the SR problem.

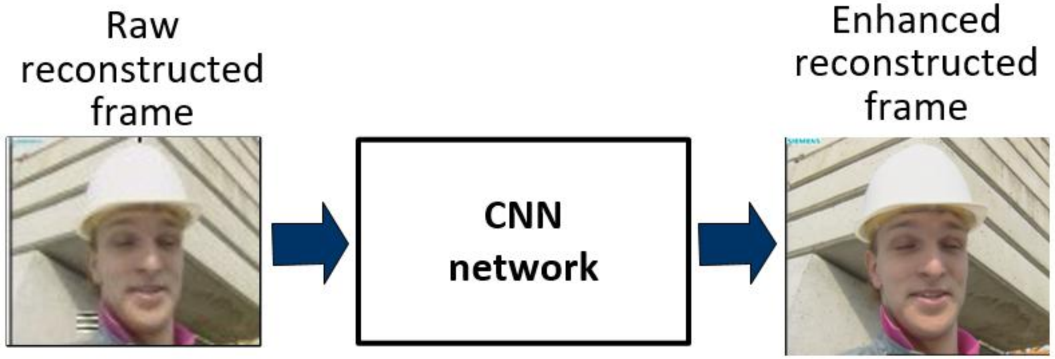


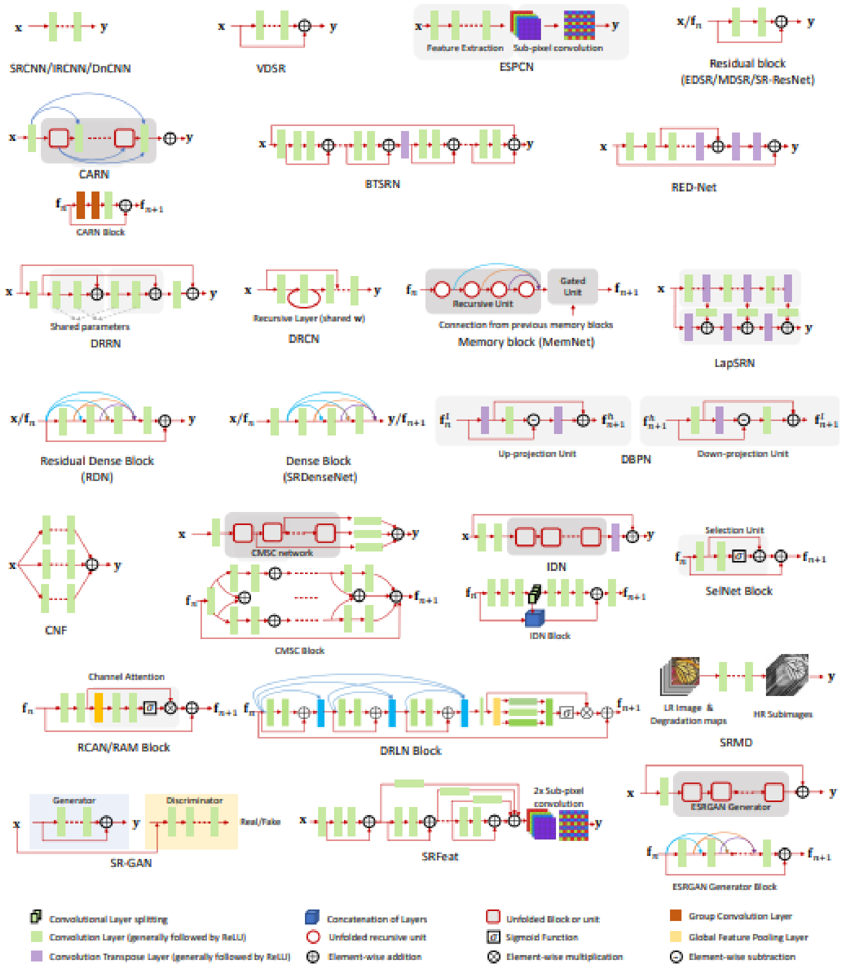


Dong et al, Learning a deep convolutional network for image super-resolution, 2014, pp. 184-199, ECCV 2014.

Loss function: $$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} ||F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i||^2$$

process the in-loop filter in the same way.



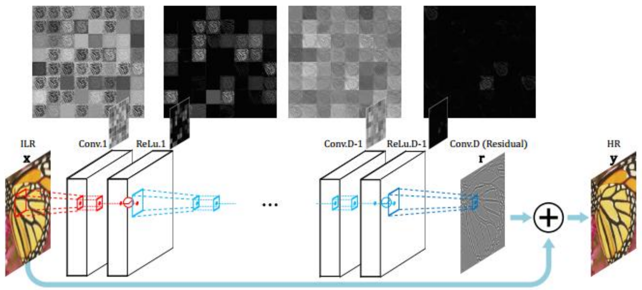


Anwar et al. A deep journey into super-resolution: A survey. Arxiv 1904.07523, 2019.

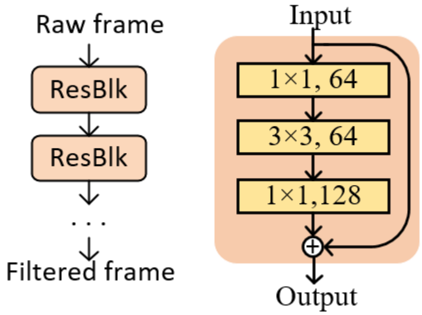
ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

Classical CNNs

VDSR

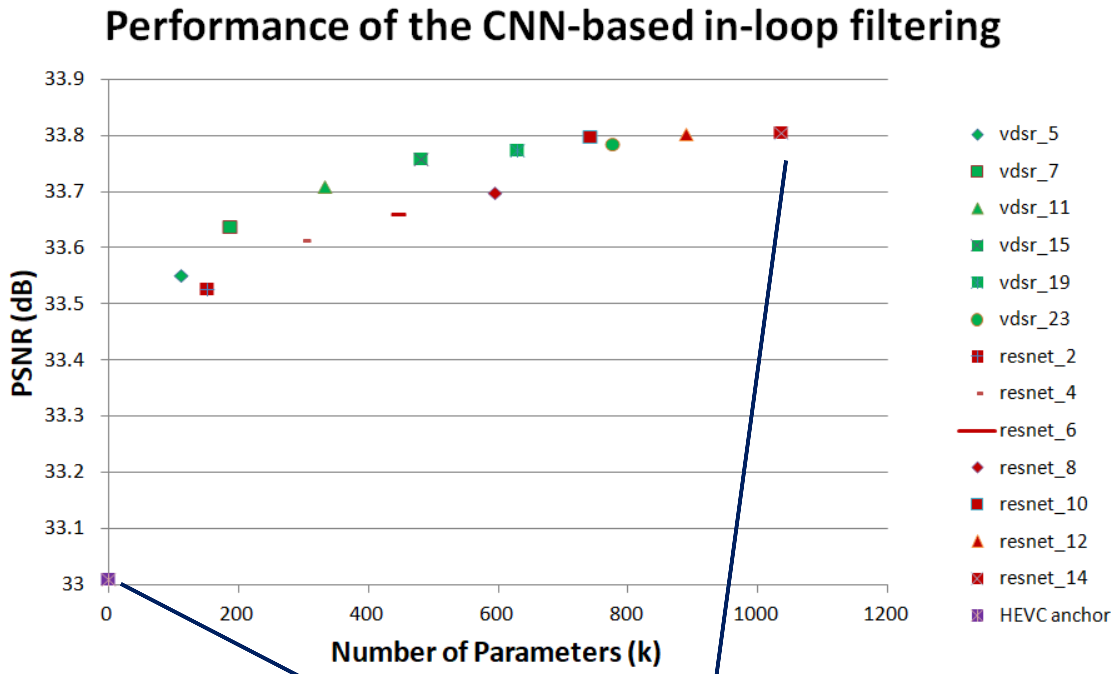


ResNet



J. Kim, et al, Accurate image super-resolution using very deep convolutional networks, pp. 1646-1654, *CVPR*, 2016.

K. He et al, Identity mappings in deep residual networks, pp. 630-645, *ECCV*, 2016.



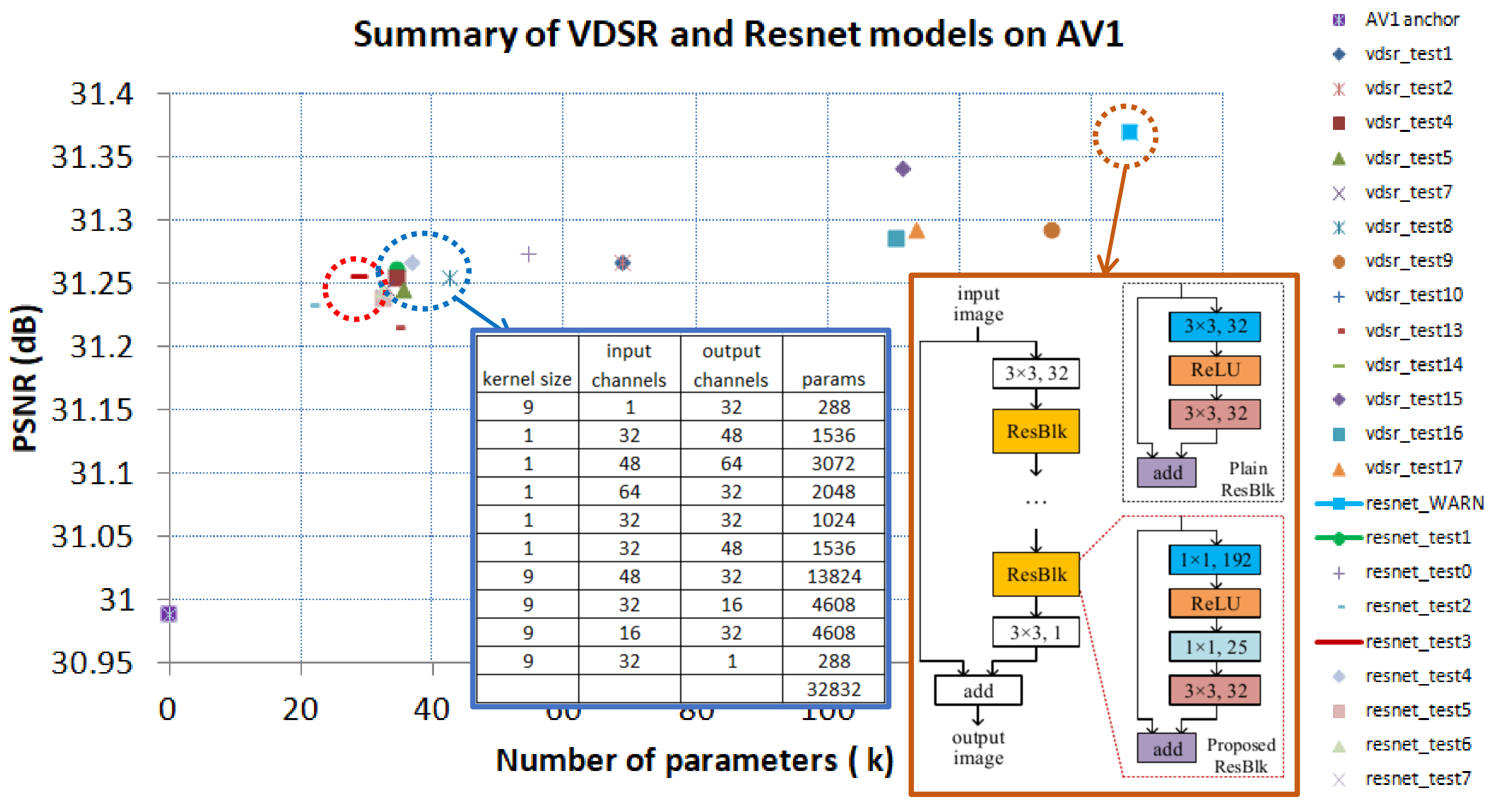
The PSNR gain is as large as 0.8dB.

Test conditions:

- HM 16.9
- 18 images
- QP=37
- Intra coding
- The anchor in-loop filters are turned off

ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

But using large amount of parameters is expensive!



kernel size	input channels	output channels	params
9	1	32	288
1	32	48	1536
1	48	64	3072
1	64	32	2048
1	32	32	1024
1	32	48	1536
9	48	32	13824
9	32	16	4608
9	16	32	4608
9	32	1	288
			32832

Test conditions

- AV1 platform (Sept.)
- 18 images
- QP=53
- Only intra coding

To obtain a slim version

- Reduces the number of channels
- Reduce the kernel size
- Select a balanced number of layers

0.25dB can be achieved with 20k parameters.

Q2

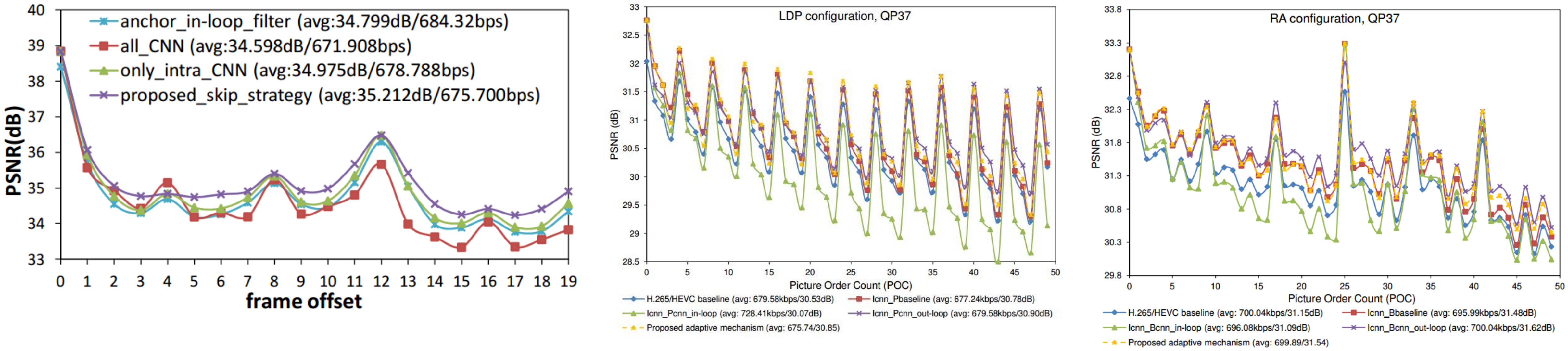
How to incorporate the CNN-based filters into video encoders?

- Previous work focuses on designing various CNN structures.
- These CNNs are directly incorporated into encoders for in-loop filtering.

Q2

How to incorporate the CNN-based filters into video encoders?

- The filtered frames will be referenced in the subsequent coding.
- Then can more gains be expected from inter coding?

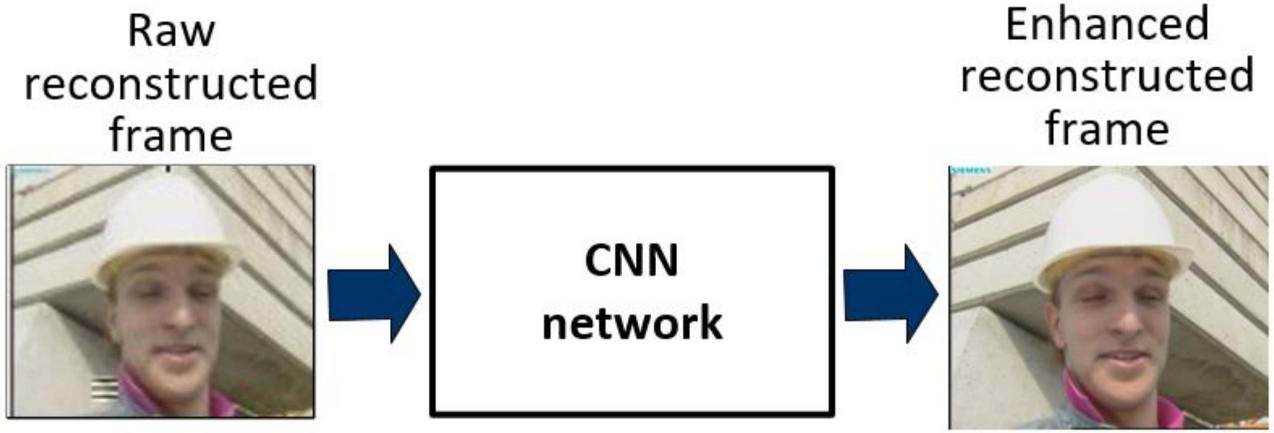


The over-filtering problem in AV1 inter (left), HEVC LDP (middle), and HEVC RA (right)

04 How to avoid the over-filtering problem?

Such a "Direct" training obtains a locally optimal model.

- A direct replacement using the "direct" model will trigger over-filtering problem.
- We cannot obtain a global optimum model because it is impossible to simulate the correlations across frame in coding.



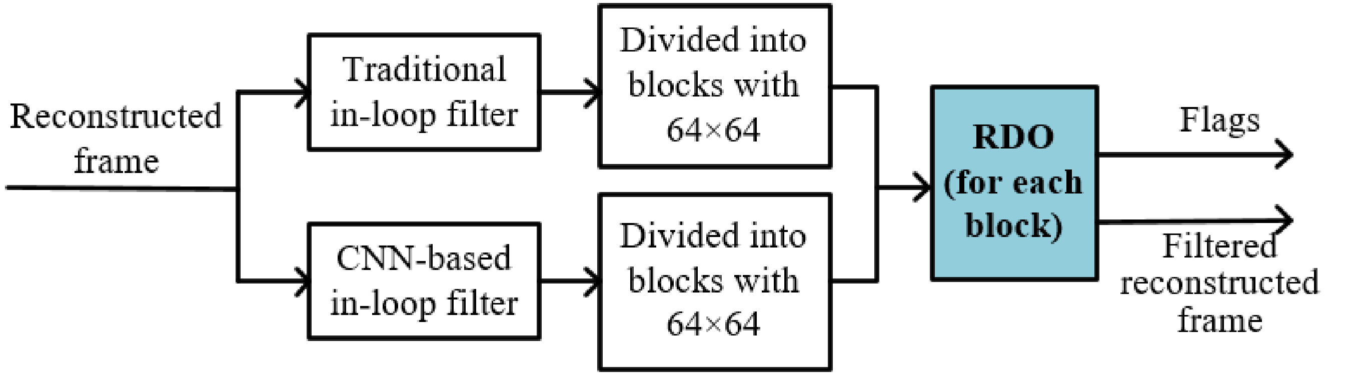
The test condition is inconsistent with the training condition.

- We conduct end-to-end training and obtain a model, without considering the intertwined correlations across frames.
- But there exists complex reference relationships in practical coding

Solution 1

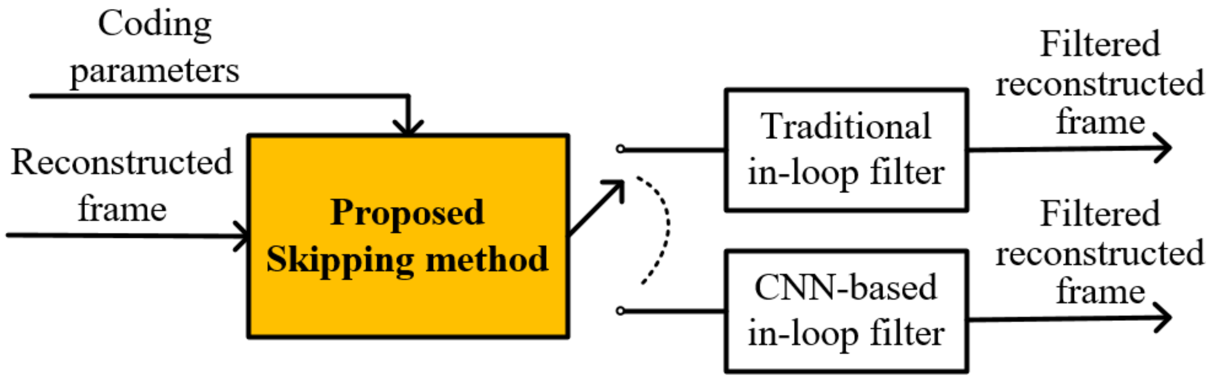
Some remedies to redress the over-filtering problem

01 Rate-Distortion method

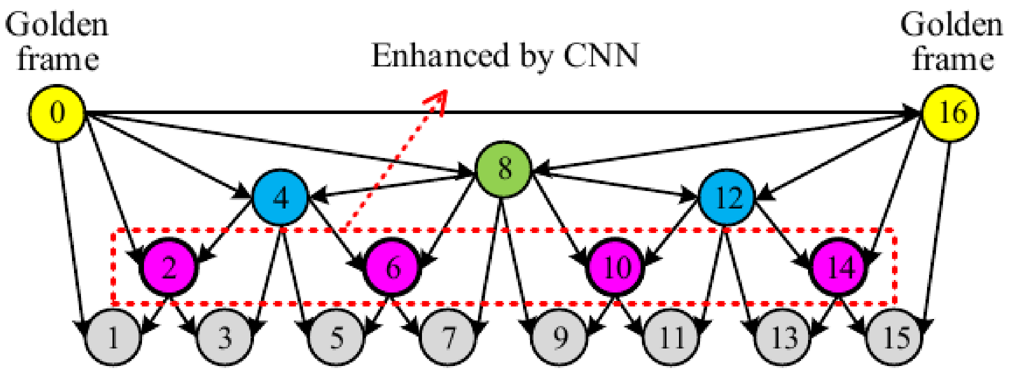


02 Skipping method

Only apply CNN to selective regions or frames



Results on AV1



Frame No.	AV1 Anchor	Enhance every frame	Proposed skipping strategy
0	31.40	31.78	31.78
1	29.80	29.68	29.96
2	29.75	29.67	29.89
3	29.33	29.01	29.45
4	29.70	29.71	29.86
5	29.14	28.95	29.37
6	29.34	29.18	29.61
7	29.27	29.08	29.57
8	29.95	30.05	30.13
Avg.	**29.74**	**29.68**	**29.96**

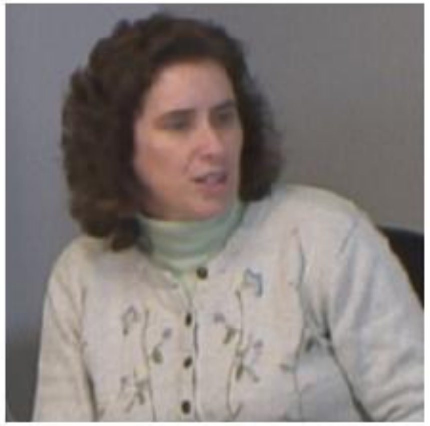
Results

- Only frame 2, 6, 10 and 14 are filtered by CNN.
- Around 0.22dB gain is retained.

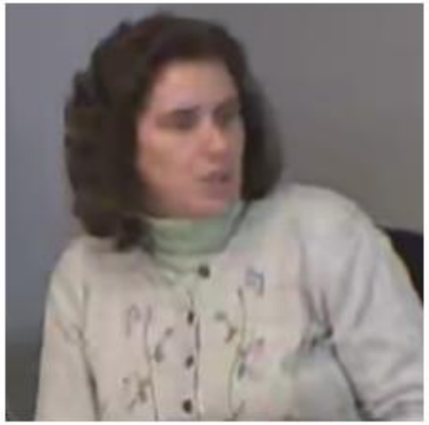
Dandan Ding, Guangyao Chen, Debargha Mukherjee, Urvang Joshi, and Yue Chen, A CNN-based in-loop filtering approach for AV1 video codec, *PCS*, 2019.

Guangyao Chen, Dandan Ding, Debargha Mukherjee, Urvang Joshi, and Yue Chen, AV1 in-loop filtering using a wide-activation structured residual network, *IEEE ICIP*, 2019.

Visual quality

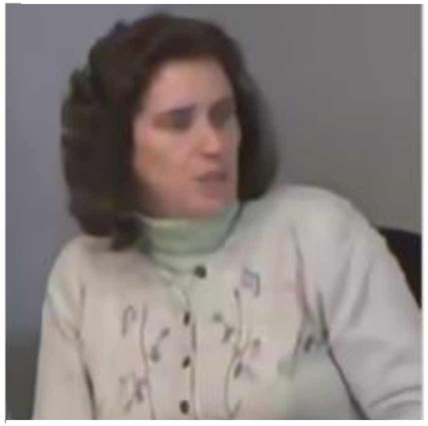


original frame

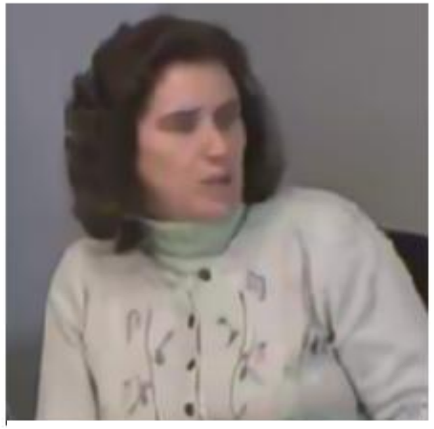


(a) Anchor

(b) Apply CNN to every frame



(c) CTU-RDO

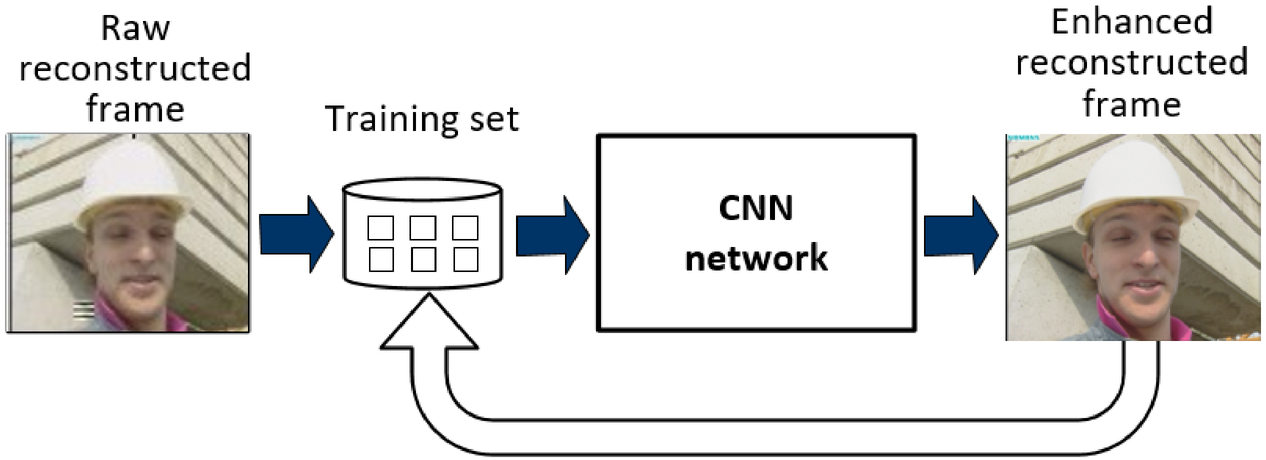


(d) Skipping method

Solution 2

Train a global model

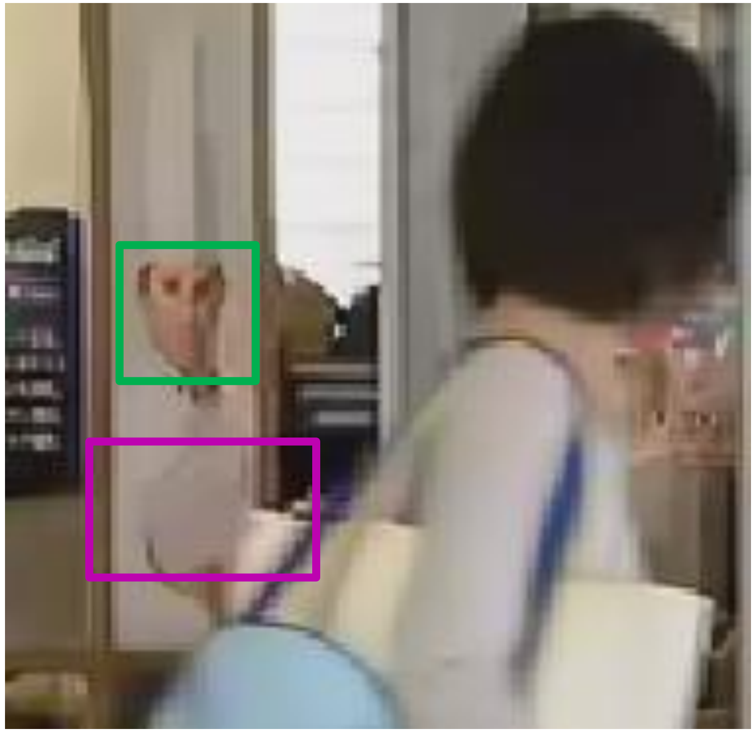
- Fundamentally solve the over-filtering problem.
- We propose a progressive training method.
  - Through transfer learning, the reconstructed frames that have been filtered by the CNN models are progressively involved back to fine-tune the CNN models themselves.



Visual quality

Original frame

CTU-RDO



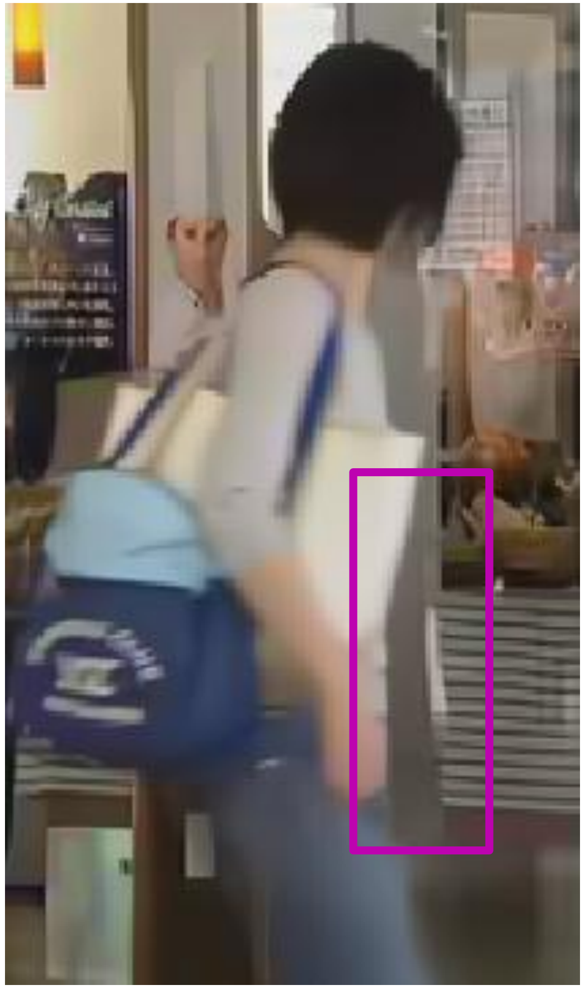
Proposed global model

ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

Original frame

CTU-RDO

Proposed global model



Results of our global model

- The global model can further improve the performance of RDO.
- A direct application of the global model to each frame will achieve a comparable gain to that of RDO.

Different solutions for over-filtering problem (PSNR)

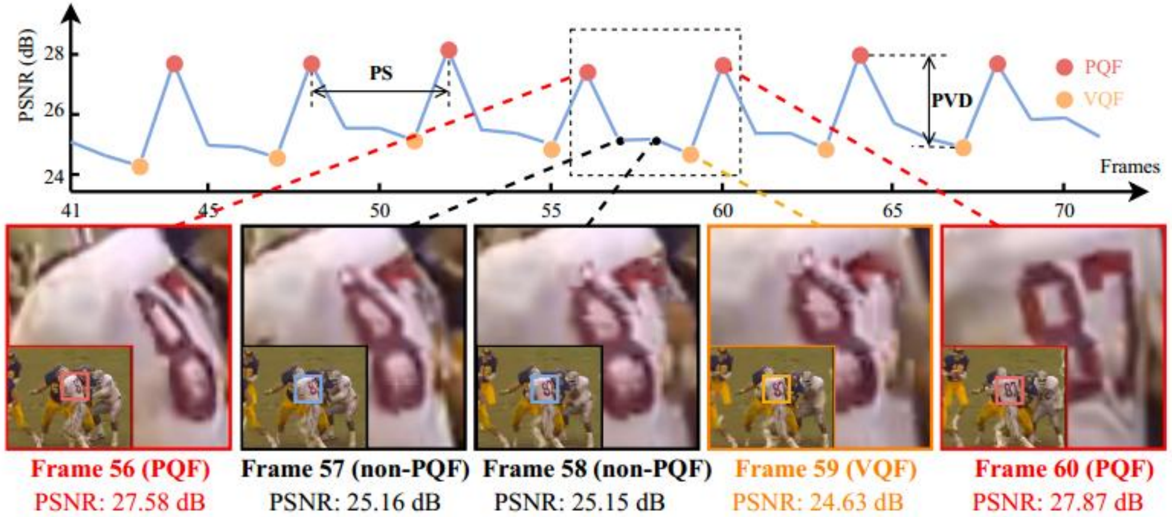
CTU-RDO using Direct model		CTU-RDO using the global model		**Directly applying the global model**	
Bitrate	PSNR	Bitrate	PSNR	**Bitrate**	**PSNR**
951.36	**32.74**	952.21	**32.79**	**942.08**	**32.77**

Test conditions

- HEVC: HM16.9
- QP=37
- 50 inter frames
- RA configuration

Multi-frame video enhancement

- Above studies are all on basis of single frame.
- Videos introduce an additional time dimension.
- How to utilize the information from temporal domain?

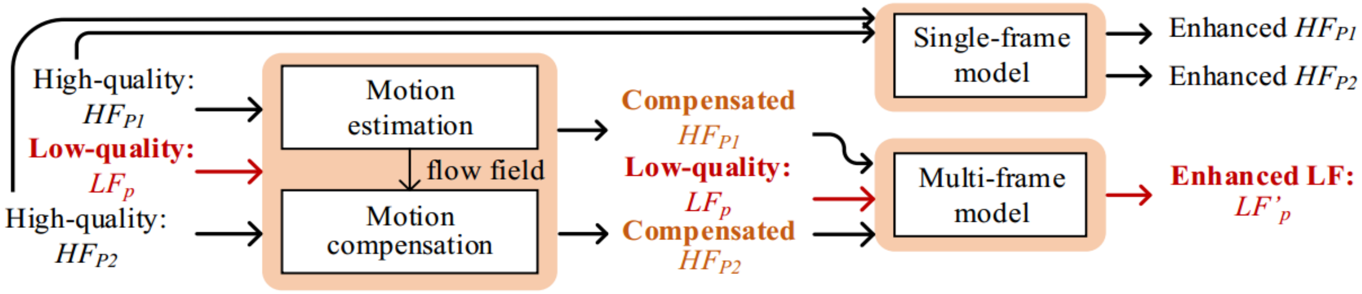


- There is frame-level quality fluctuation in compressed videos.

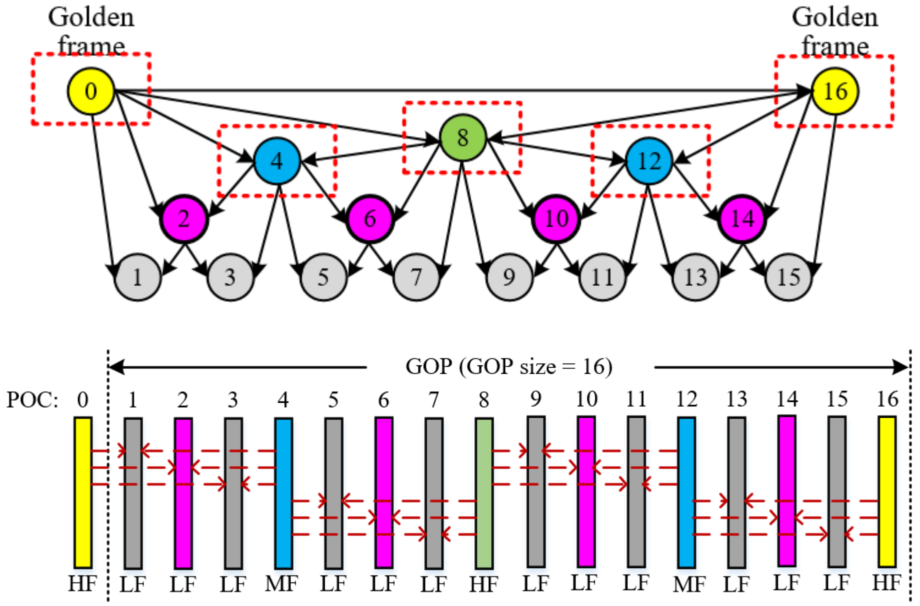
- A pair of high-quality frames can be utilized to enhance the low-quality frames in between.

R. Yang, et al, Multi-frame quality enhancement for compressed video," pp. 6664-6673, 2018, *CVPR*, 2018.

Results on AV1



Dandan Ding, Zheng Zhu, and Zoe Liu, Learning-based multi-frame video quality Enhancement, *IEEE ICIP*, 2019.



Performance of multi-frame method on AV1 (PSNR)

	AV1 anchor	**Multi frame**	Single frame
Inter	35.49	**35.84**	35.71
Intra	30.72	**31.57**	30.86

Test conditions

- QP=53
- Only 36 low-quality frames
- Flownet2.0 is employed for motion estimation

Conclusion

- Two problems are concerned when embedding the CNN-based tools into video encoders.
  - The CNN structure
  - The incorporation approaches
- Currently, we employ a single CNN model to deal with all videos.
- It is possible to develop different small CNNs for different video characteristics.

Thank You

DandanDing@hznu.edu.cn

https://github.com/IVC-Projects