Improved Deep Image Compression with Joint Optimization of Cross Channel Context Model And Generalized Loop Filter

Changyue Ma, Zhao Wang, Ru-Ling Liao, Yan Ye

Alibaba Group

Contents

1. Introduction

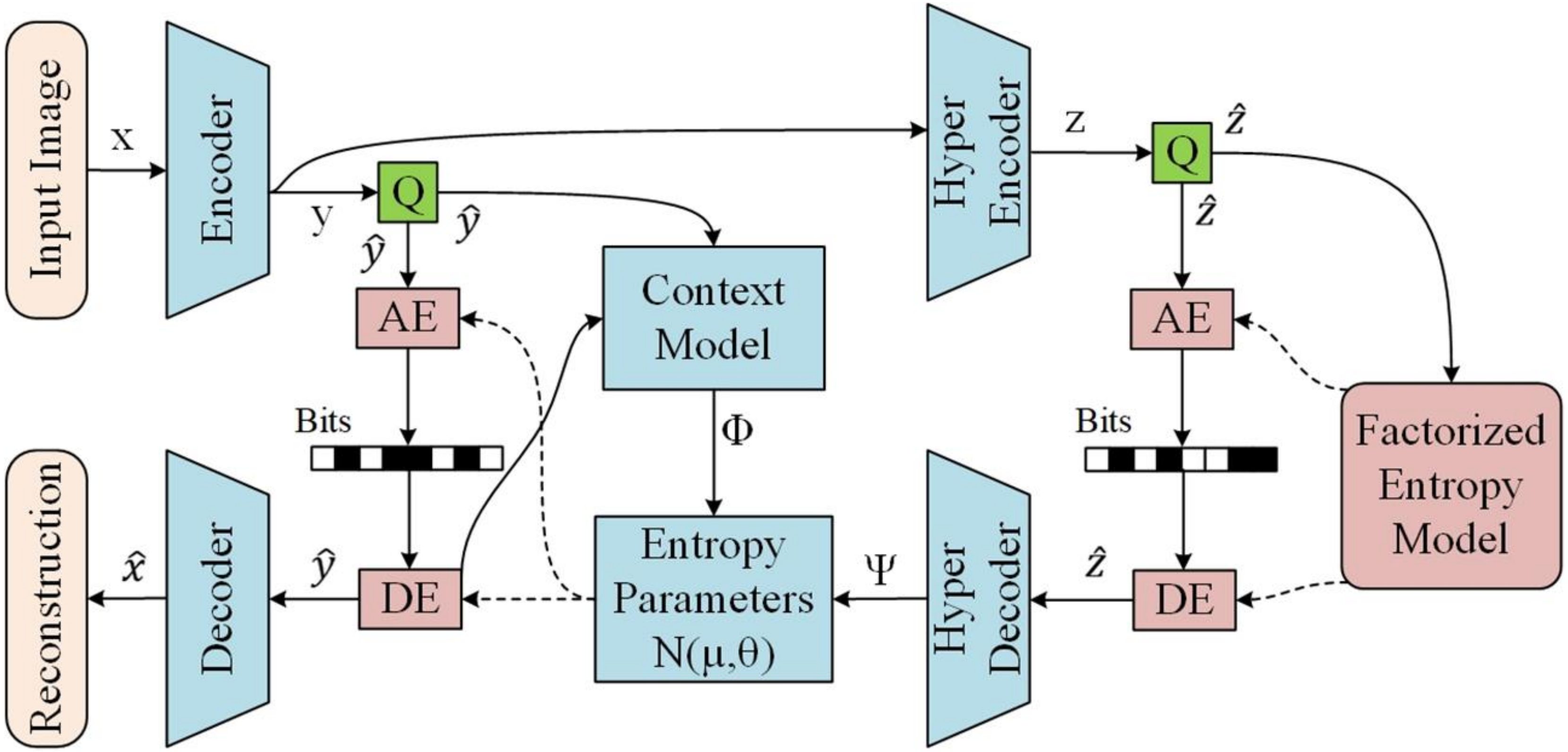


Figure 1. Block diagram of the joint autoregressive and hierarchical prior model for deep image compression.

1. Introduction

Coding performance improvement   Our Aim

1) Make all steps in the deep image compression framework differentiable.
2) Introduce Gaussian mixture model with its parameters estimated by combining the hyperprior and an autoregressive mask convolution.
3) Introduce the residual connection and the attention module in network structure of transform.

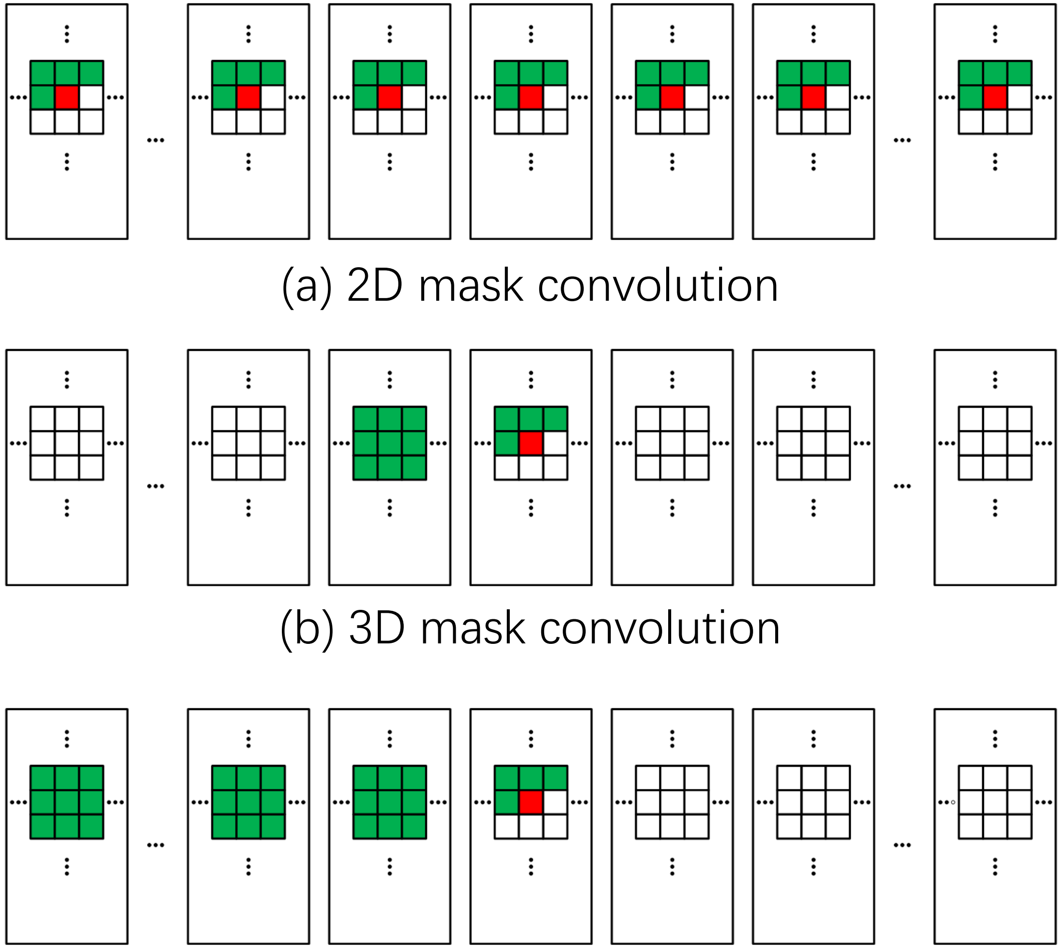
Computational complexity reduction

1) Divide the quantized latents into several oblique slices and encode them one by one.
2) Use the channel-wise context model to replace the spatial context model.

2. Proposed Methods

- Cross channel context model
- Generalized loop filter
- Implementation details

2.1 Cross Channel Context Model

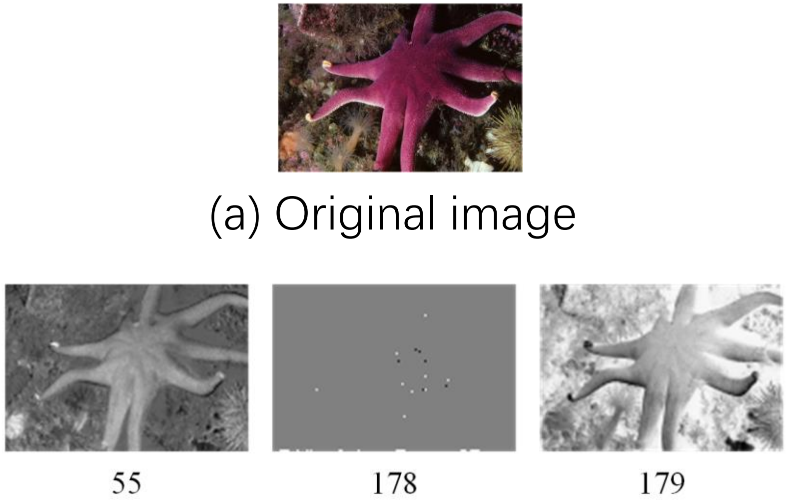


(a) 2D mask convolution

(b) 3D mask convolution

(c) The proposed cross channel context model

Fig.2 Different context models for latents in deep image compression.

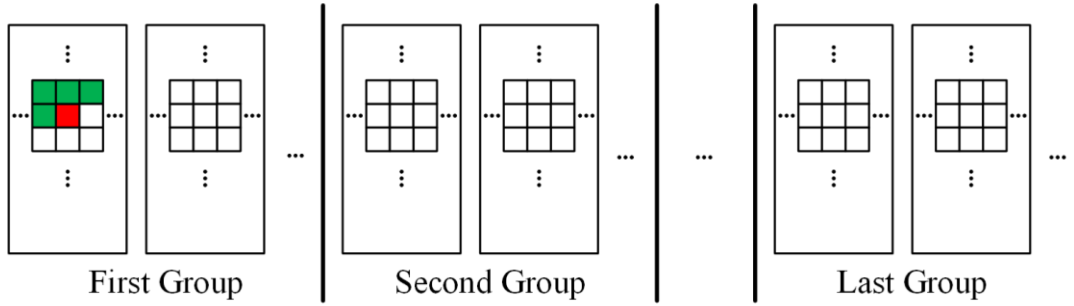


(a) Original image

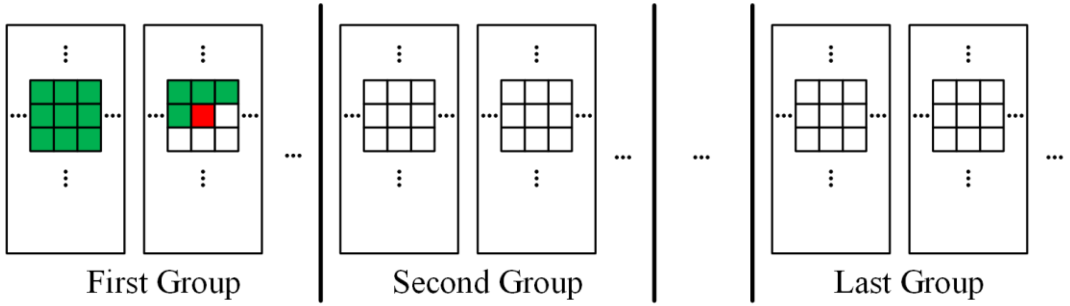
(b) Example of correlation comparison

Fig.3 Cross channel correlation analysis.

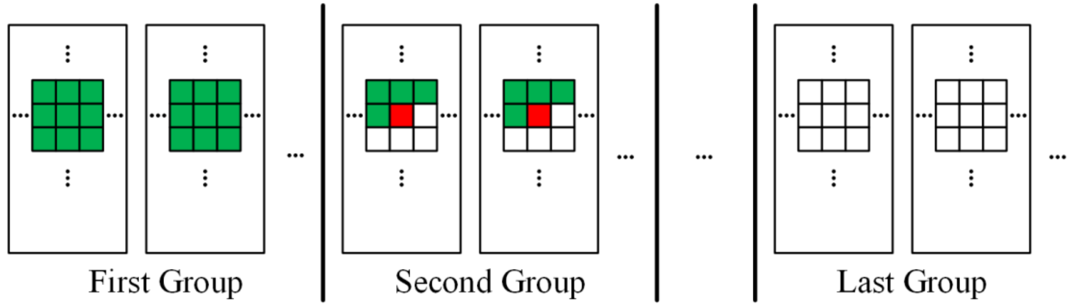
2.1 Cross Channel Context Model



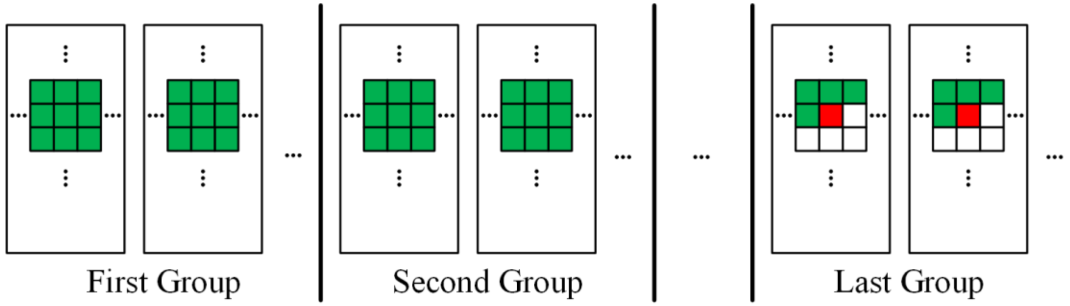
(a) Coding the first channel in the first group.



(b) Coding other channels in the first group.



(c) Coding the second group.



(d) Coding the last group.

Figure 4. Basic steps of the proposed cross channel context model.

2.2 Generalized Loop Filter

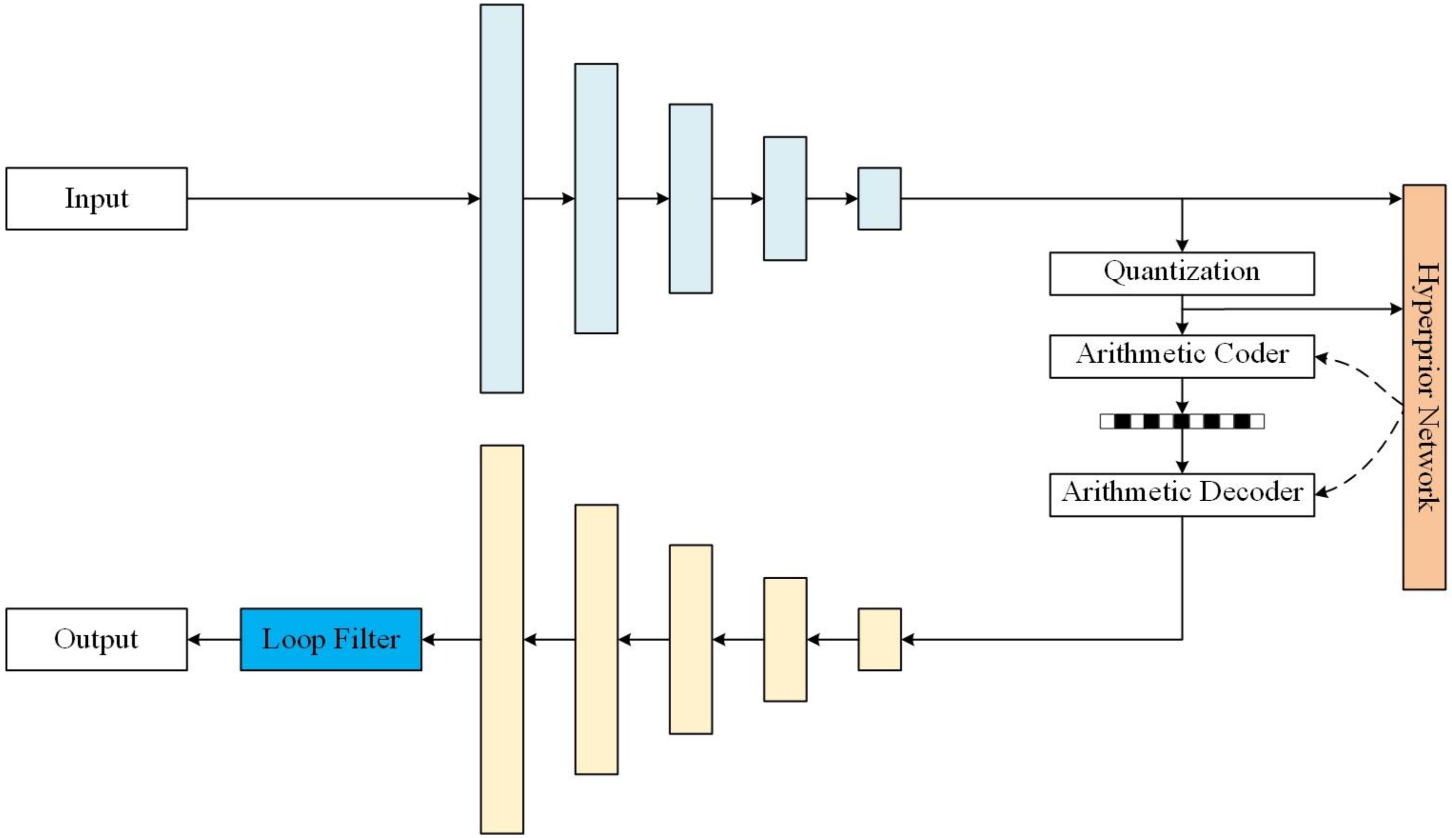


Figure 5. Basic loop filter and deep image compression combination scheme.

2.2 Generalized Loop Filter

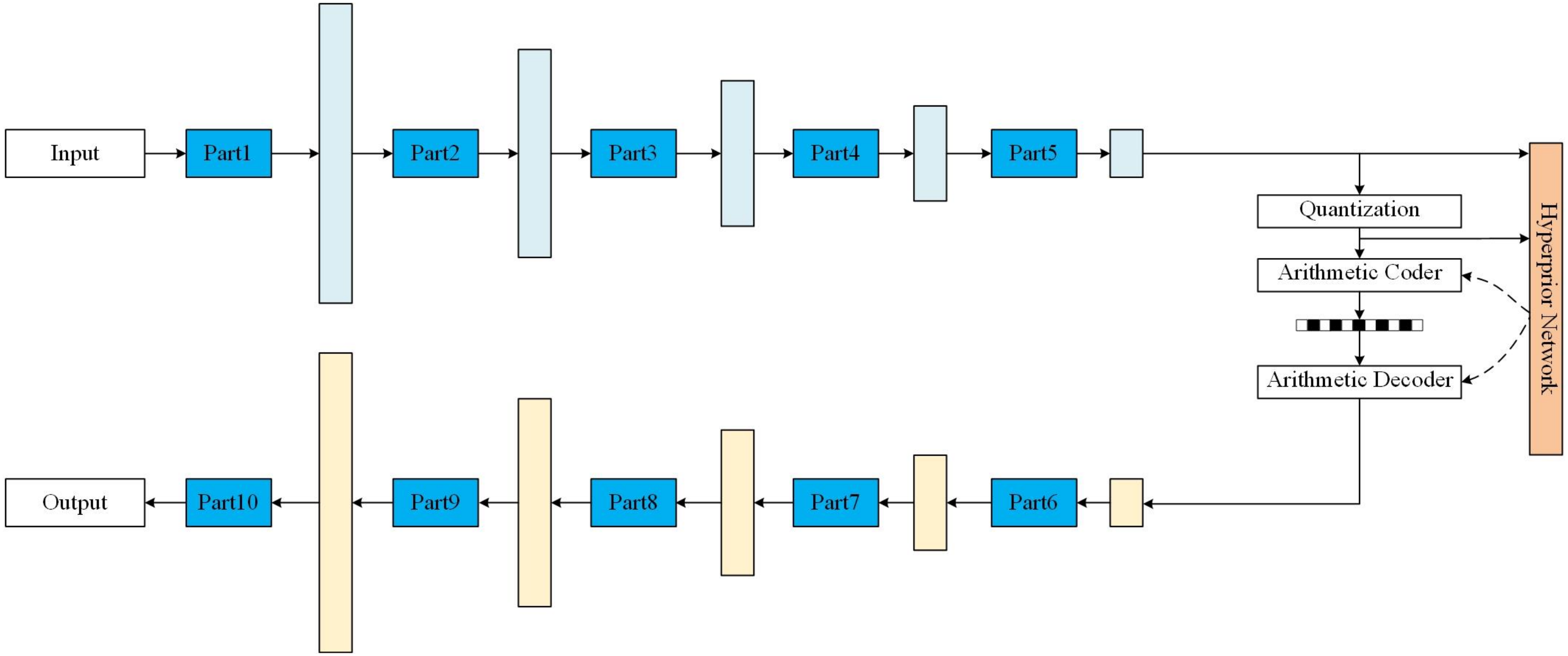


Figure 6. Generalized loop filter and deep intra frame compression combination scheme.

With limited number of basic network units, we attempt:

1) placing all the basic network units in the forward transform.
2) placing all the basic network units in the inverse transform.
3) equally placing the network units in both the forward and inverse transforms. (Best)

2.3 Implementation Details

1) We implement the proposed methods into the *cheng2020-attn* model in CompressAI.
2) For cross channel context model, the latents are uniformly divided into 8 groups according to the channel index.
3) For generalized loop filter, we equally insert 3 residual blocks into both forward and inverse transform.
4) The original deep image compression model designed to compress images in RGB domain is adjusted to compress the images in YUV420 domain.

3. Training and Testing Details

Training Details:

Dataset: DIV2K, UCID
Patch Size: 256x256(Y), 128x128(UV)
Iterations: 1M
Batch Size: 16
Optimizer: Adam
Learning Rate: 1e-4 for 0.7M iterations, then divided by 2
Distortion Weight: MSE YUV 4:1:1

Testing Details:

Dataset: Kodak (converted to YUV420 format)
Class A1, A2, B, C, and E in JVET CTC sequences

4. Experimental Results

Sequences		*cheng2020-attn*			*cheng2020-attn* + CCCM + GLF		
		Y	U	V	Y	U	V
JVET CTC	Class A1	8.90%	87.73%	40.89%	−0.18%	67.96%	21.30%
	Class A2	1.35%	47.53%	46.07%	−7.18%	33.49%	31.46%
	Class B	8.10%	30.10%	25.75%	−0.48%	15.70%	11.09%
	Class C	13.55%	23.24%	26.64%	1.34%	4.05%	4.16%
	Class E	9.26%	36.07%	4.62%	−2.27%	15.40%	−5.34%
JVET CTC Average		**8.51%**	**42.08%**	**28.34%**	**−1.44%**	**24.74%**	**11.91%**
Kodak Average		**7.20%**	**2.85%**	**7.48%**	**−1.20%**	**−10.82%**	**−5.38%**

Table 1. BD-rate reduction of *cheng2020-attn* and *cheng2020-attn* integrated with the proposed cross channel context model (CCCM) and generalized loop filter (GLF) over VTM-11.0.

(1) There is coding performance loss when comparing original *cheng2020-attn* model with VTM11.0.
(2) The coding performance of deep image compression outperforms (Kodak) or is close to (JVET CTC) VTM11.0 after integrating the proposed CCCM and GLF methods.
(3) The coding performance of deep image compression on Y, U and V components can be adjusted through adjusting the training distortion weight for Y, U and V components.

4. Experimental Results

Sequences		*cheng2020-attn* + CCCM			*cheng2020-attn* + CCCM + GLF		
		Y	U	V	Y	U	V
JVET CTC	Class A1	−4.38%	−3.32%	−8.04%	−8.17%	−10.08%	−15.09%
	Class A2	−1.05%	−3.45%	−3.89%	−8.24%	−11.10%	−11.23%
	Class B	−3.71%	−5.77%	−8.74%	−7.72%	−10.58%	−11.76%
	Class C	−6.05%	−11.52%	−11.41%	−10.92%	−15.23%	−16.48%
	Class E	−5.75%	−12.05%	−0.49%	−10.73%	−14.50%	−8.16%
JVET CTC Average		**−4.24%**	**−7.30%**	**−7.03%**	**−9.10%**	**−12.27%**	**−12.68%**
Kodak Average		**−4.45%**	**−7.44%**	**−7.92%**	**−7.80%**	**−12.66%**	**−11.15%**

Table 2. BD-rate reduction of *cheng2020-attn* integrated with the proposed cross channel context model (CCCM) and generalized loop filter (GLF) over *cheng2020-attn*.

(1) The proposed CCCM method brings around 4% coding performance improvements.
(2) The proposed CCCM and GLF methods bring around 7%~9% coding performance improvements.

5. Conclusion

(1) We proposed a cross channel context model and generalized loop filter to further improve the coding performance of deep image compression.

(2) We adjusted the original deep image compression model designed to compress images in RGB domain to compress the images in YUV420 domain.

(3) With the proposed methods, the coding performance of deep image compression outperforms (Kodak) or is close to (JVET CTC) VVC when measured with PSNR distortion in YUV420 domain, which to our knowledge is the first such achievement.

(4) The proposed methods are applicable in both image compression and intra coding in video compression.

Thank you!