Learned Image Compression for Both Humans and Machines via Dynamic Adaptation

Lingyu Zhu[1], Binzhe Li[1], Riyu Lu[1], Peilin Chen[1],
Qi Mao[3], Zhao Wang[4], Wenhan Yang[5] and Shiqi Wang[1,2]

- [1] City University of Hong Kong
- [2] Shenzhen Research Institute (CityU)
- [3] Communication University of China
- [4] Peking University
- [5]PengCheng Laboratory

Email: lingyzhu-c@my.cityu.edu.hk

Introduction

Background

- Increasing Data Volume
  - Rapid development of multimedia applications has led to a massive increase in image and video data.
  - Efficient compression of this data is a fundamental challenge in multimedia communication and processing.

- Human vs. Machine Vision
  - Human Vision: Requires realistic and visually pleasing signals with rich appearances and textures.
  - Machine Vision: Focuses on restoring rich semantic clues for analytics tasks.
  - The difference necessitates dedicated compression methods for each.

Introduction

Motivation

- Challenges in Optimization
  - Techniques optimized for human perception may reduce machine analysis performance.
  - Increasing focus on compression for machines to enhance performance in tasks like detection and segmentation.
  - Need for joint optimization of machine vision tasks under bitrate constraints.

- Innovative Approaches
  - Dynamic adaptation of representations to align with task-driven requirements.
  - Aim to achieve higher compression ratios without significant loss of semantic information, facilitating both image reconstruction and machine vision tasks.

Framework

香港城市大學
City University of Hong Kong

Dynamic Adaptation for Humans and Machines

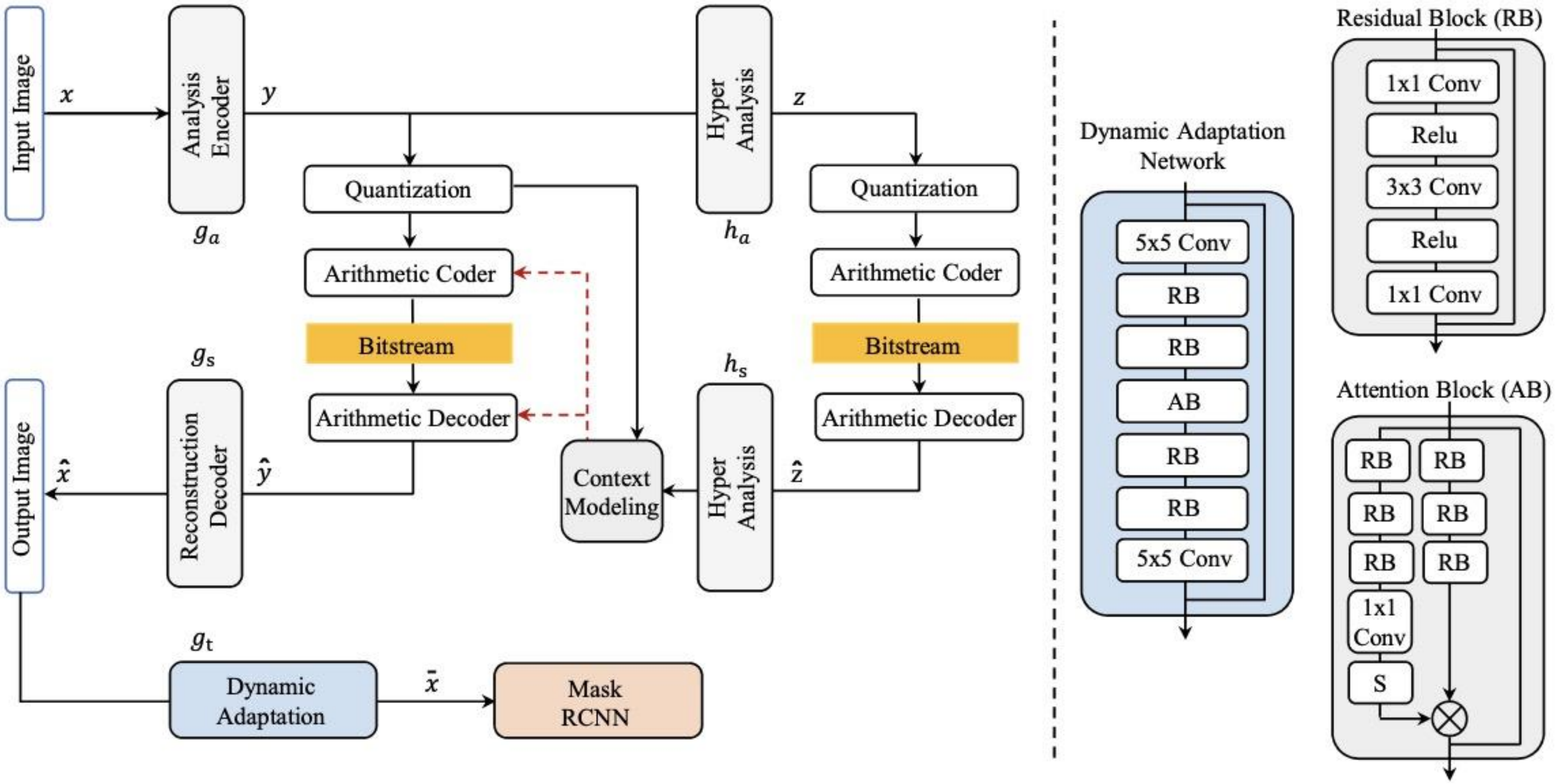
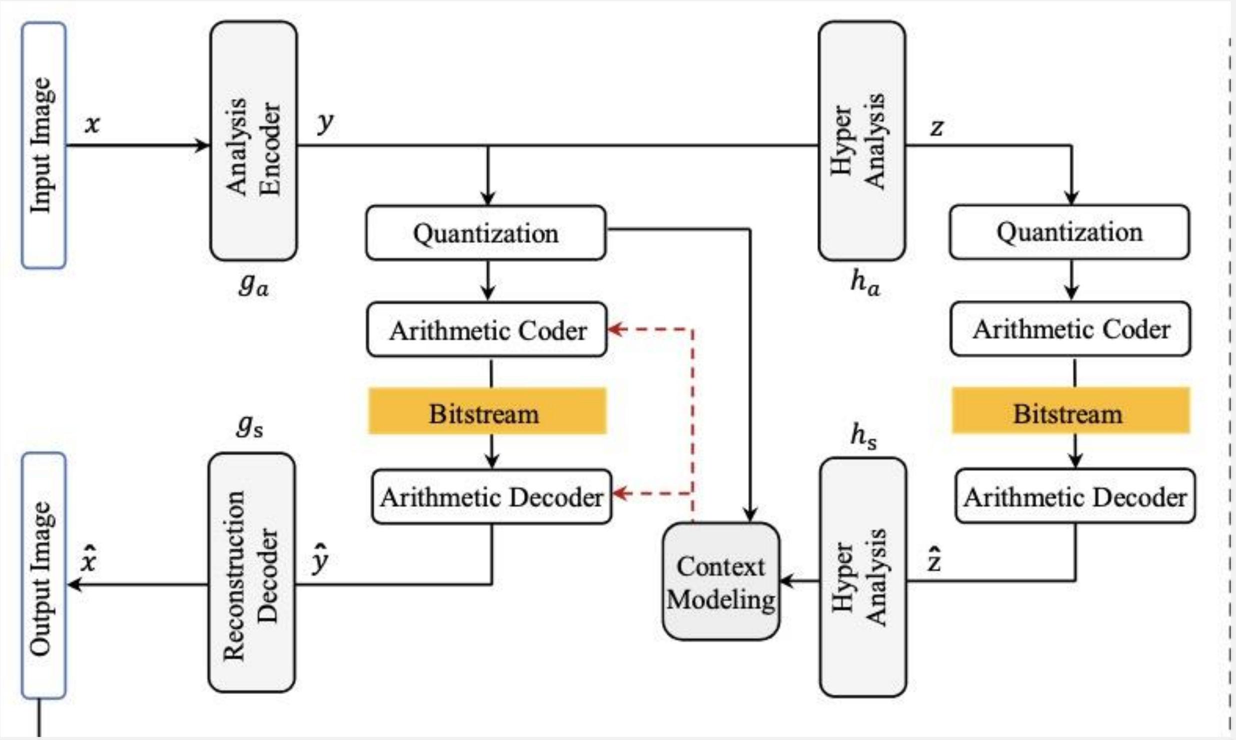


Figure 1. Overall architecture of the proposed method.

Framework

Human Perception Oriented Compression



Rate-distortion Trade-off

- **Rate Calculation**

$$R = E\left[-\log_2 p_{\hat{y}}(\hat{y})\right] + E[-\log_2 p_{\hat{z}}(\hat{z})]$$

Latent Hyper-prior

- **Distortion Calculation**

$$D_h = MSE(x, \hat{x})$$

Figure 2. Overall architecture of the proposed method.

Framework

Machine Analysis Oriented Adaptation

Distribution to Machine Vision

$$D_m = L_{cls}^{c} + L_{reg} + L_{cls}^{p} + L_{loc} + L_{mask}$$

Classifier loss — Proposal loss — Mask loss

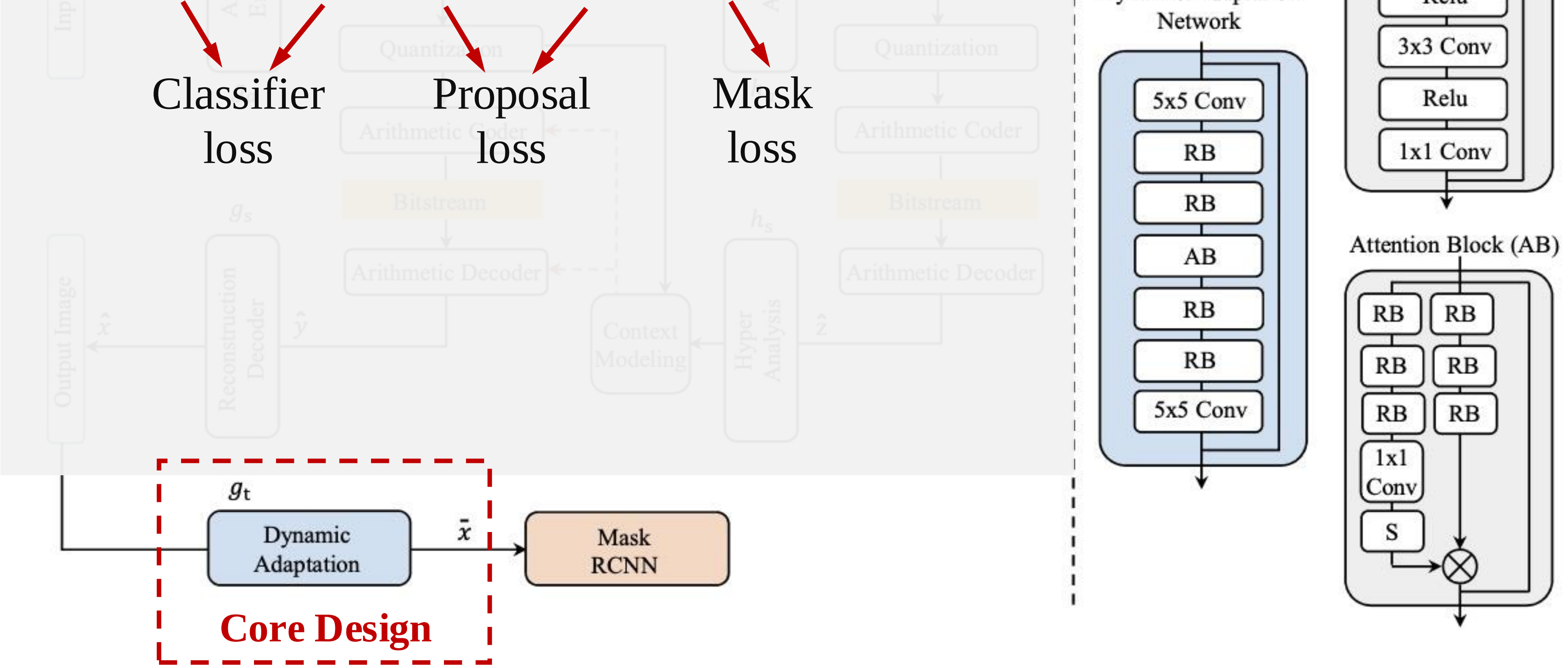


Figure 3. Overall architecture of the proposed method.

Framework

Machine Analysis Oriented Adaptation

Distribution to Machine Vision

$$D_m = L_{cls}^{c} + L_{reg} + L_{cls}^{p} + L_{loc} + L_{mask}$$

Classifier loss

Proposal loss

Mask loss

Total Optimization Loss

$$L_{rdo} = MSE(\hat{y}, \hat{z}) + \lambda_h D_h(x, \hat{x}) + \lambda_m D_m(x, \bar{x})$$

Rate term

Distortion term

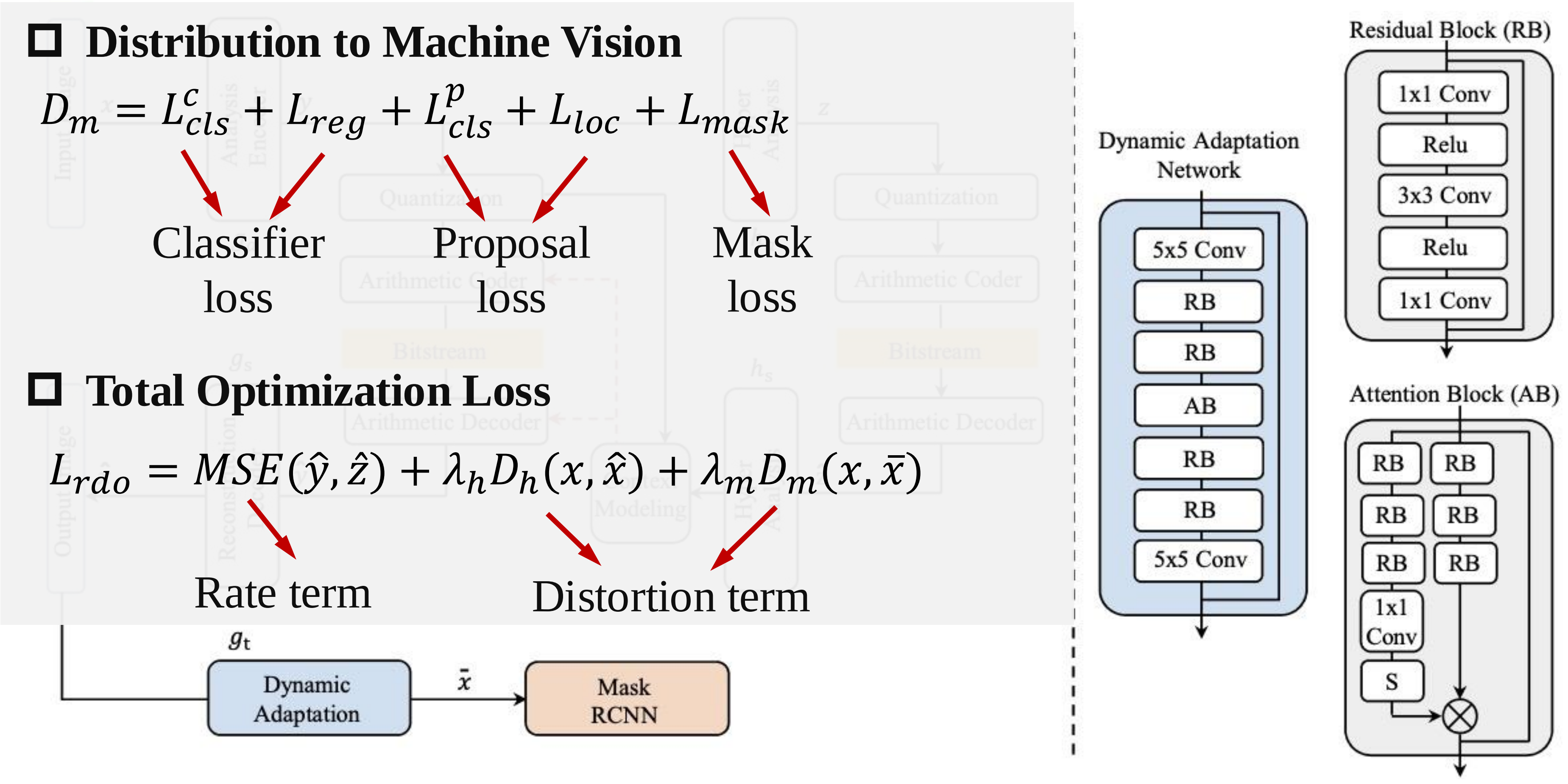


Figure 4. Overall architecture of the proposed method.

Experimental Details

- Training Dataset:
  - COCO 2017 training set
  - 118,287 natural images
  - Commonly used for object detection and segmentation.

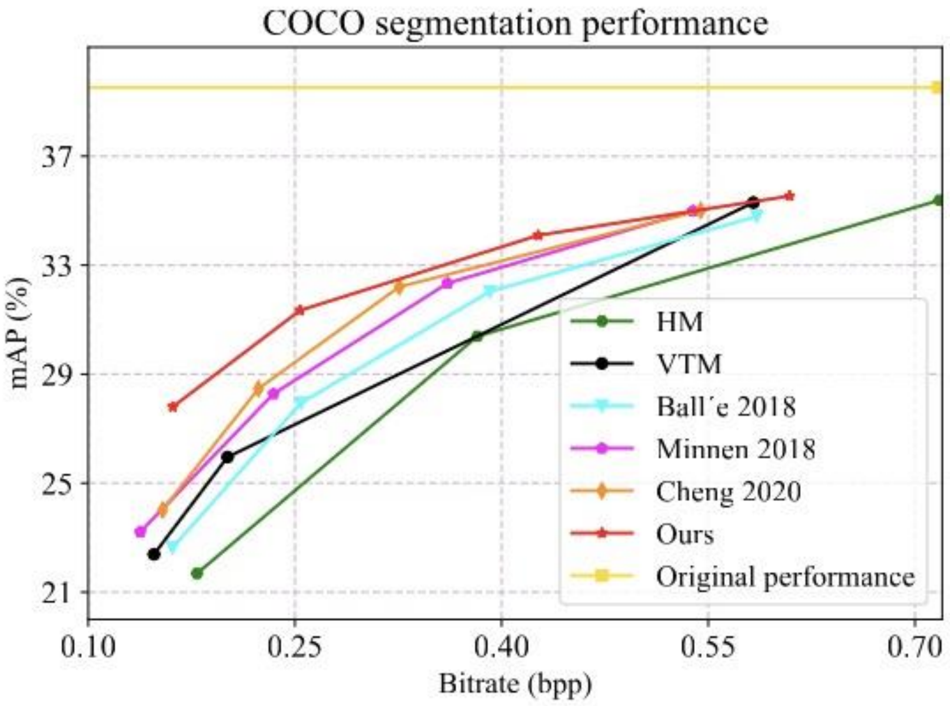
- Testing Data:
  - Machine Vision: COCO 2017 Validation Dataset (5,000 images).
  - Human Vision: Kodak Dataset (24 high-quality images).

- Performance Evaluation:
  - Machine Vision: Mean Average Precision (mAP) with IoU threshold from 0.5 to 0.95 (interval 0.05).
  - Human Vision: BD-Rate savings based on PSNR.

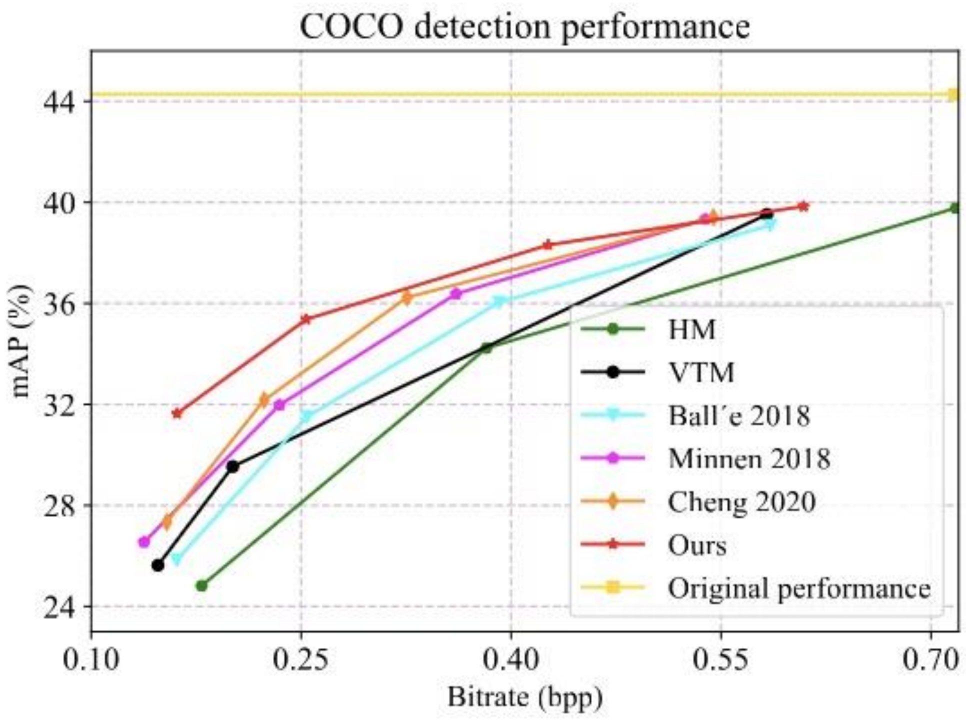
Experimental Results

- Achieve superior performance in both the detection and segmentation tasks

☐ Segmentation performance



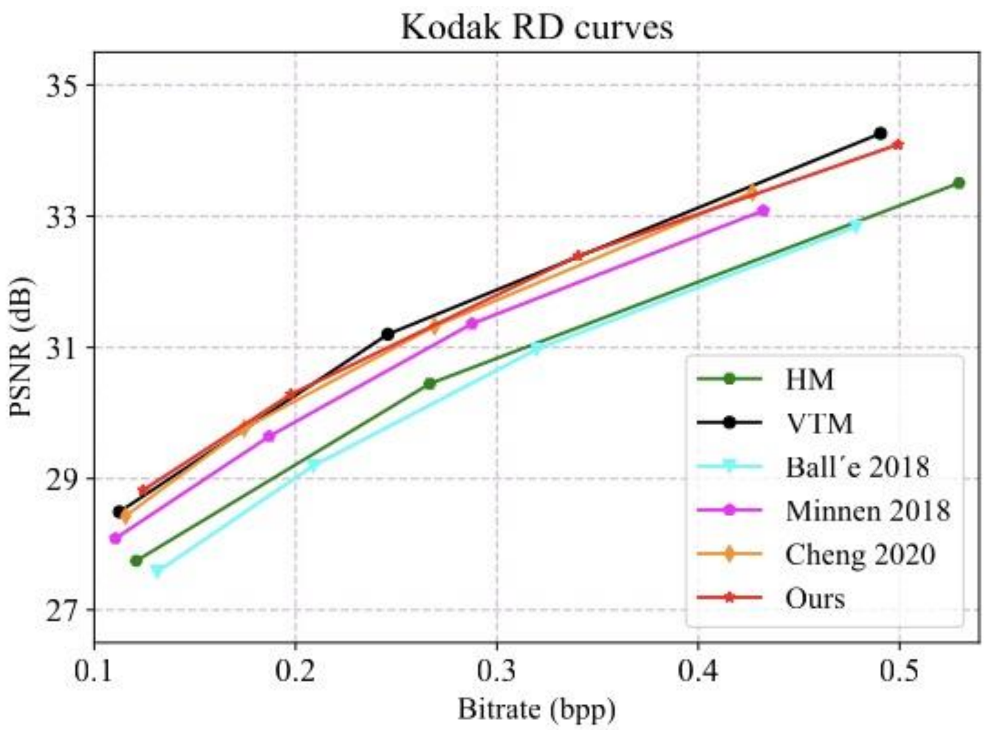
☐ Detection performance



Experimental Results

- Achieve promising performance in reconstruction

Human-vision performance



Overall performance

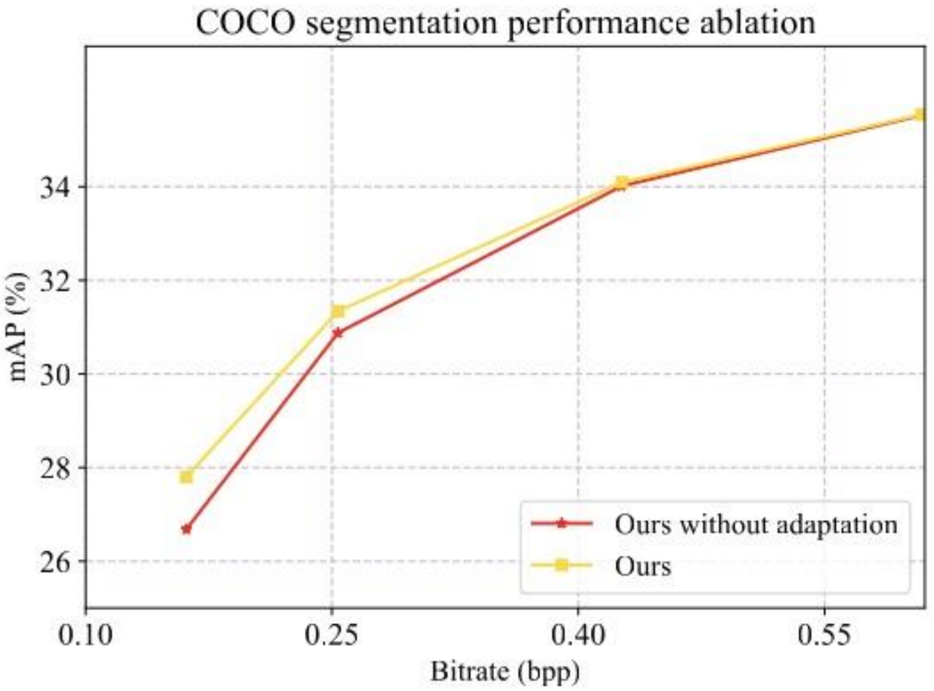
Table 1. BD-rate and BD-mAP of the proposed method for comparison. Herein, the HEVC is the anchor.

Benchmarks	COCO BD-mAP (Detection)	COCO BD-mAP (Segmentation)	Kodak BD-rate (PSNR)
HEVC Intra [5]	-	-	-
VVC Intra [7]	-23.78%	-23.98%	-25.20%
Ball´e 2018 [14]	-19.08%	-18.37%	+7.68%
Minnen 2018 [16]	-28.79%	-27.89%	-13.85%
Cheng 2020 [2]	-31.52%	-30.77%	-19.76%
Ours	-36.76%	-36.79%	-21.50%

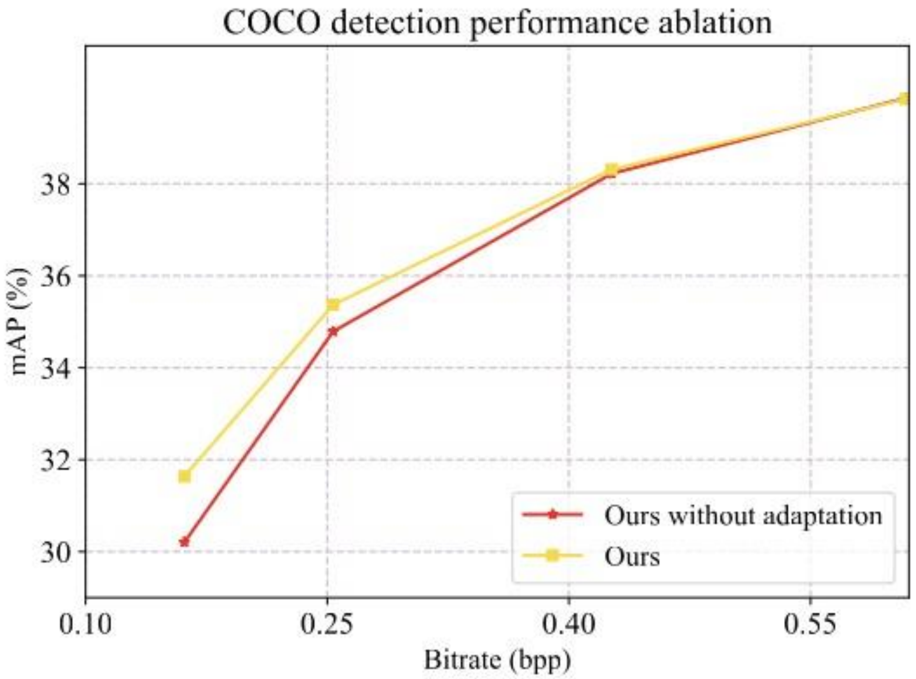
Ablation Results

• Achieve a positive impact at the low bitrate in experimental observation.

☐ Segmentation performance



☐ Detection performance



Conclusion

- Dynamic Adaptation Approach: The proposed method successfully integrates human and machine vision through a dynamic transformation network that adjusts data distribution.

- Rate-Distortion Performance: Improved performance metrics are achieved for both human and machine vision, indicating a significant advancement in image processing.

- End-to-End Optimization: The optimization of the dynamic adaptation network enhances its applicability across various image datasets, showcasing the method's versatility.

Thank you!

Link to Github

Link to Paper