Analysis on Compressed Domain: A Multi-Task Learning Approach

Yuefeng Zhang[1], Chuanmin Jia[1], Jianhui Chang[1], Siwei Ma[1,2,3]

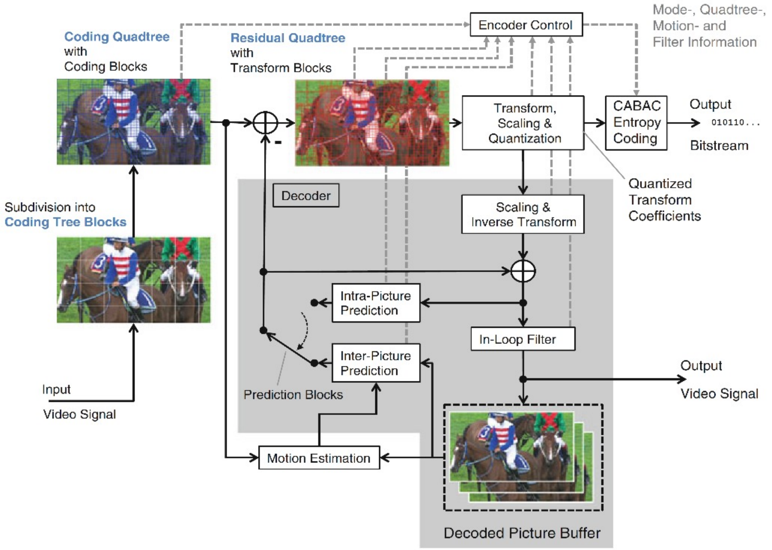
1. Peking University, 2. Peng Cheng Laboratory,

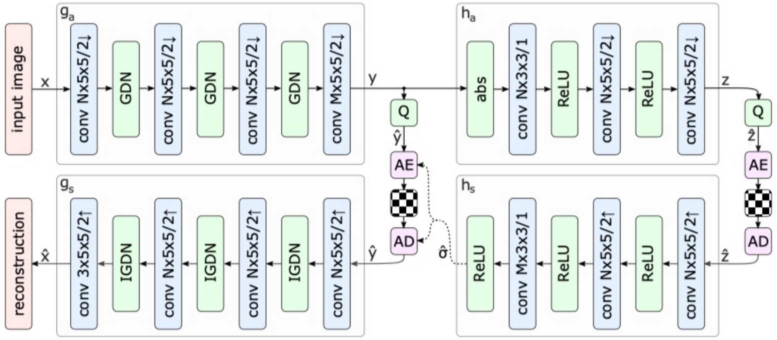
3. Information Technology R&D Innovation Center of Peking University

2022.3

Data Compression Conference

Two Problems into One





Problem 1: Image Compression

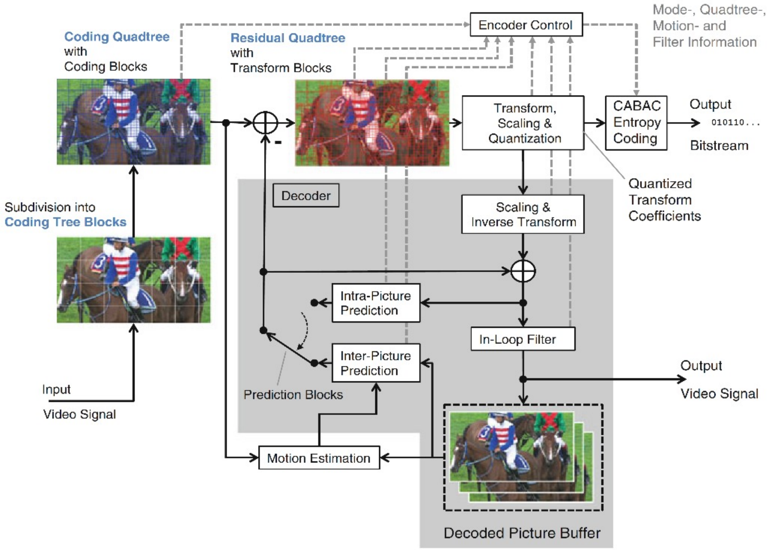
Image Classification

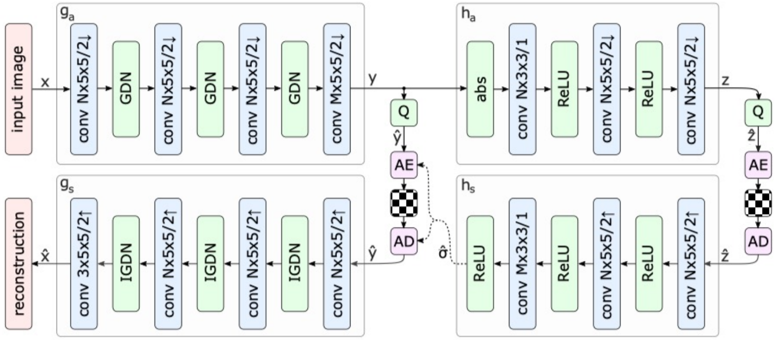
Face Recognition

Semantic Segmentation

Problem 2: Image Perception

Two Problems into One





Problem 1: Image Compression

&

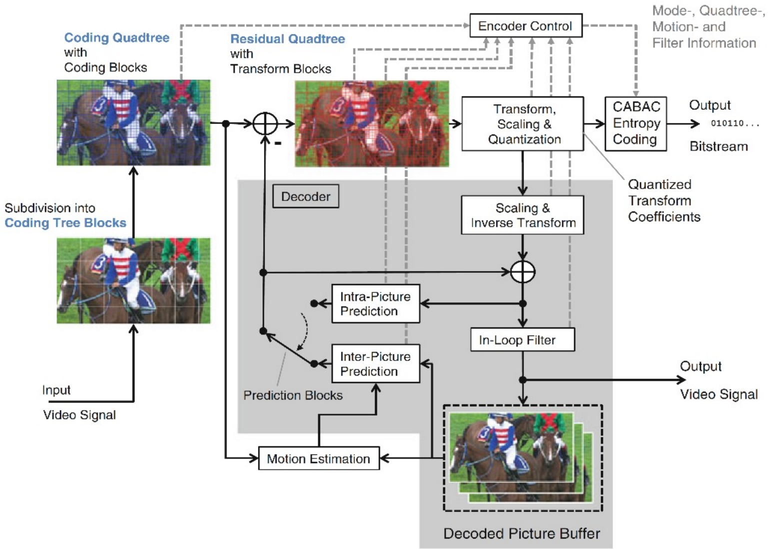
Image Classification

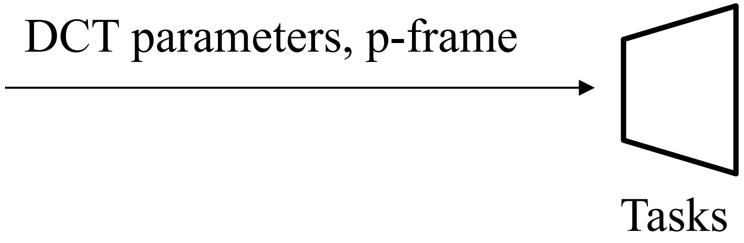
Face Recognition

Semantic Segmentation

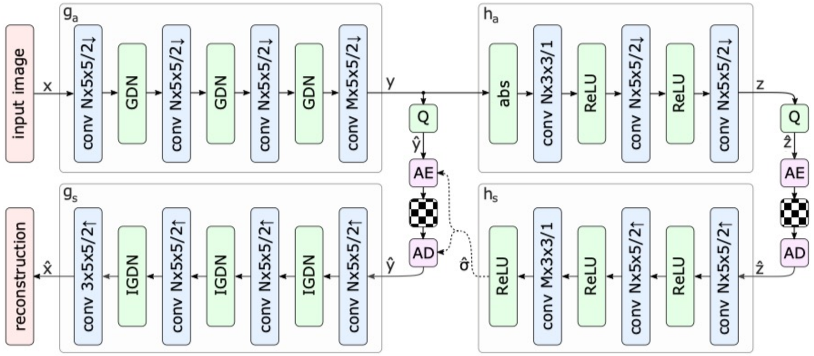
Problem 2: Image Perception

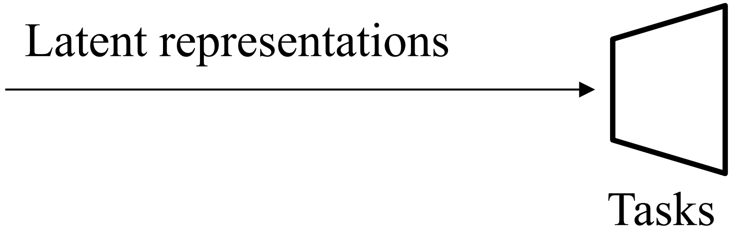
Related Work





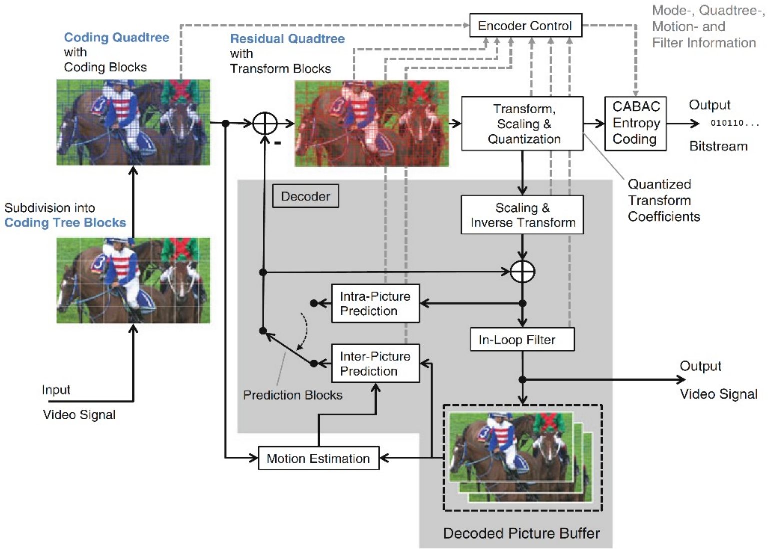
Gueguen et al. 2018
Ehrlich et al. 2019
Wu et al. 2018



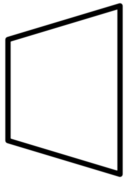


Alvar et al. 2019
Choi et al. 2021

Related Work



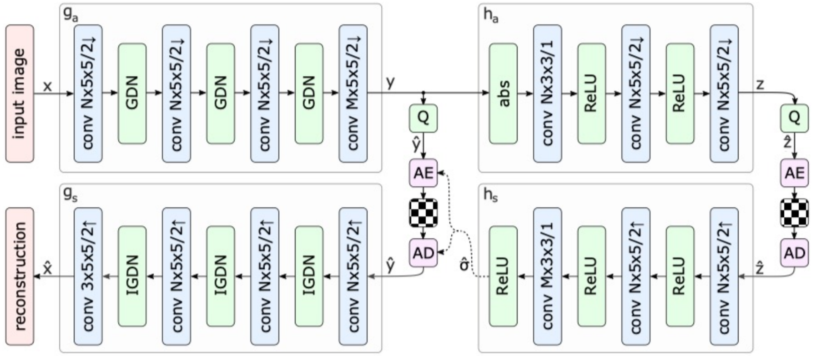
DCT parameters, p-frame



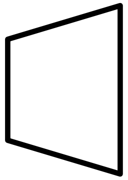
Tasks

Gueguen et al. 2018
Ehrlich et al. 2019
Wu et al. 2018

Lack of learning ability.



Latent representations

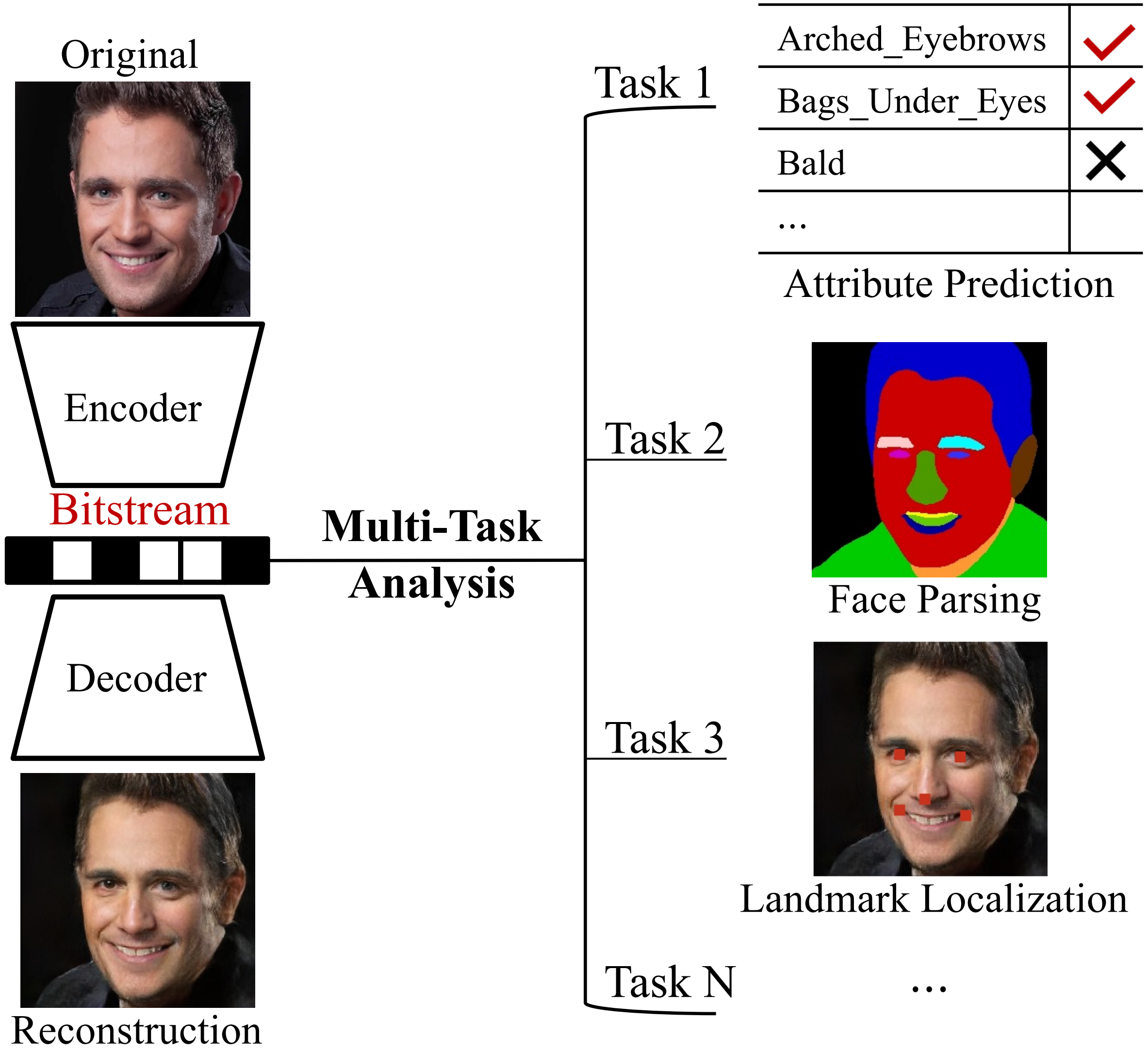


Tasks

Alvar et al. 2019
Choi et al. 2021

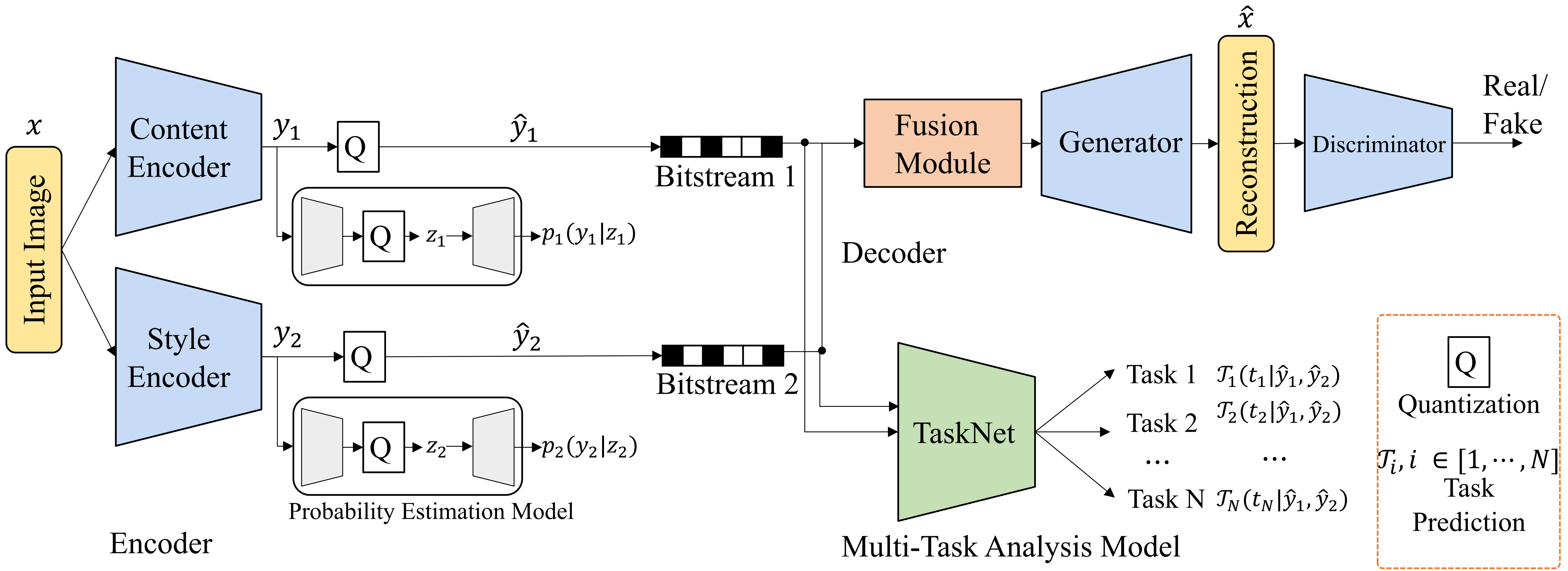
Dynamic application scenes are challenging.

Proposed Method

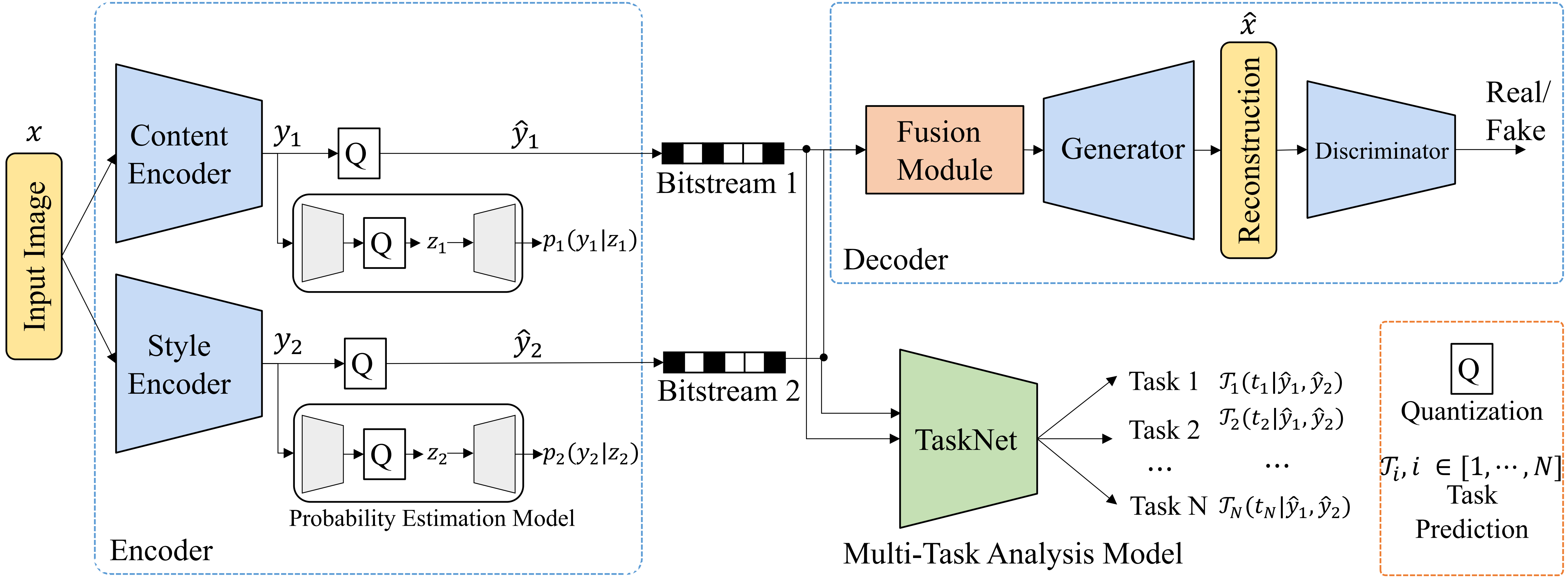


- **End-to-end** training.
- Taking **latent representations** as downstream visual tasks' input.
- One model supports **multiple** visual tasks.

Proposed End-to-end Optimization Model

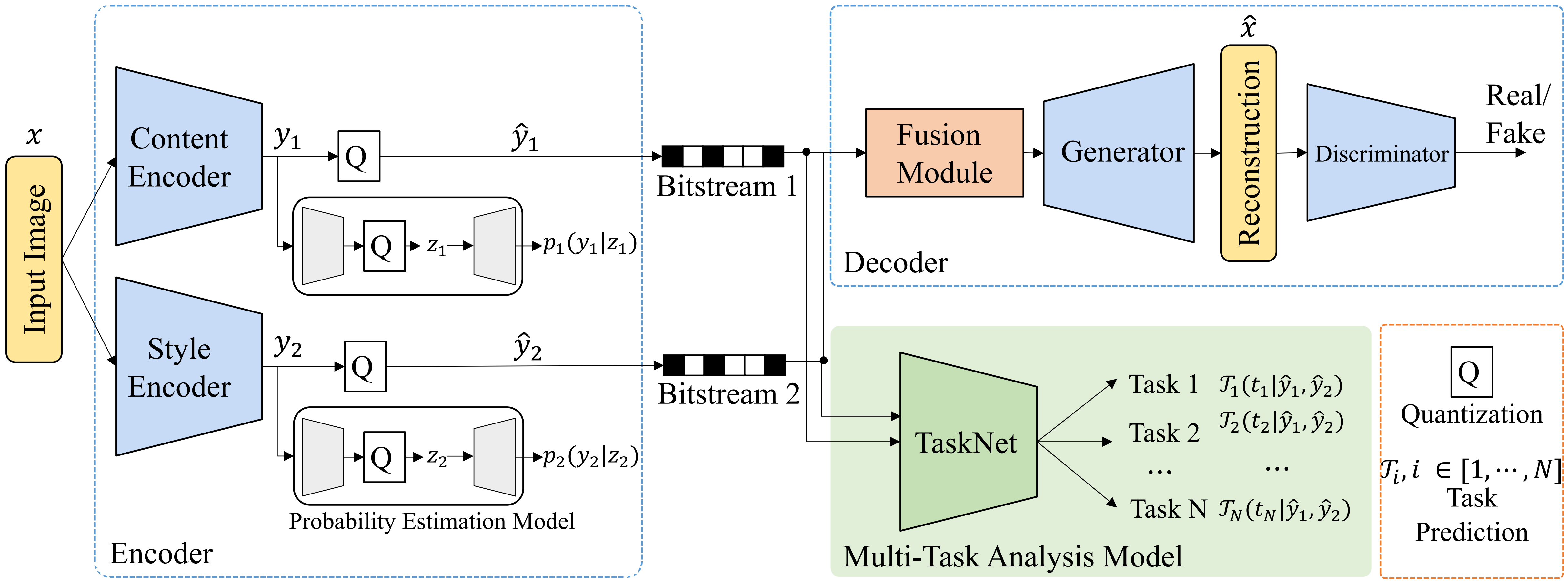


Proposed End-to-end Optimization Model



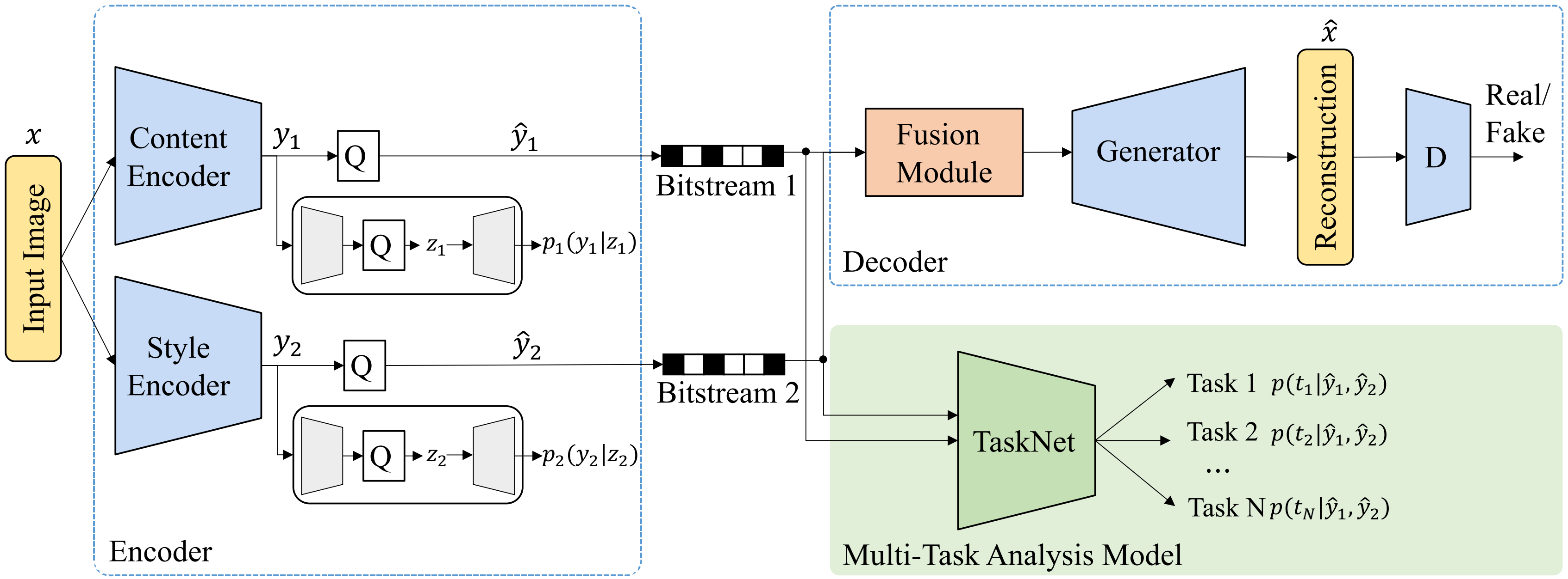
- **Layered** representations.
- **Adversarial** learning.

Proposed End-to-end Optimization Model



- Treat downstream visual tasks in a **multiple** learning way.
- Trade-off among **rate, distortion and analysis efficiency**.

Proposed End-to-end Optimization Model



Encode & Decode：

$$y = E(x), y' = Q(y), x' = D(y')$$

Distortion:

$$d = k_{MAE} d_{MAE} + k_{SSIM} d_{SSIM} + k_p d_p$$

Adversarial Loss:

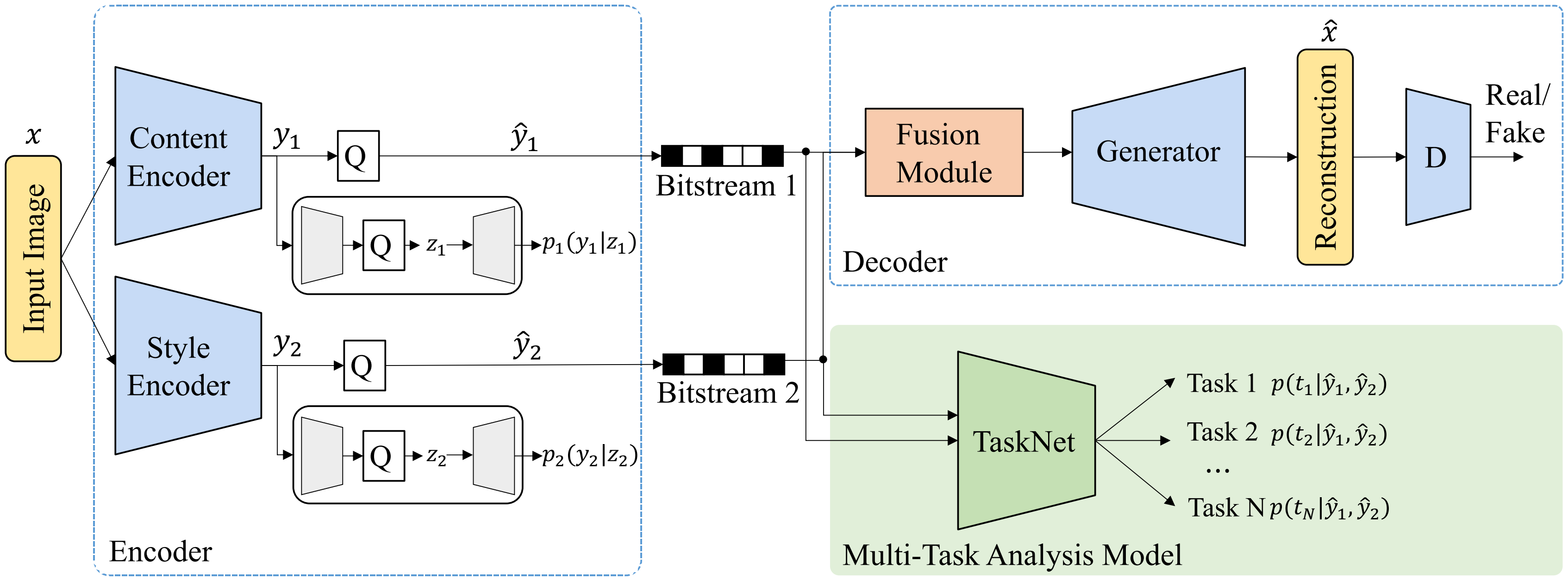
$$\mathcal{L}_{EGP} = \mathbb{E}_{x \sim p_X}[-\lambda \log(P(y')) + d(x, x') - \beta \log(D(x', y'))]$$

$$\mathcal{L}_D = \mathbb{E}_{x \sim p_X}[-\log(1 - D(x', y'))] + \mathbb{E}_{x \sim p_X}[-\log(D(x, y'))]$$

Analysis Loss:

$$\mathcal{L}_{multi} = \lambda_{cls} l_{cls} + \lambda_{seg} l_{seg}$$

Proposed End-to-end Optimization Model



Encode & Decode:

$$y = E(x), y' = Q(y), x' = D(y')$$

Distortion:

$$d = k_{MAE}d_{MAE} + k_{SSIM}d_{SSIM} + k_p d_p$$

Adversarial Loss:

$$\mathcal{L}_{EGP} = \mathbb{E}_{x\sim p_X}[-\lambda\log(P(y')) + d(x,x') - \beta\log(D(x',y'))]$$

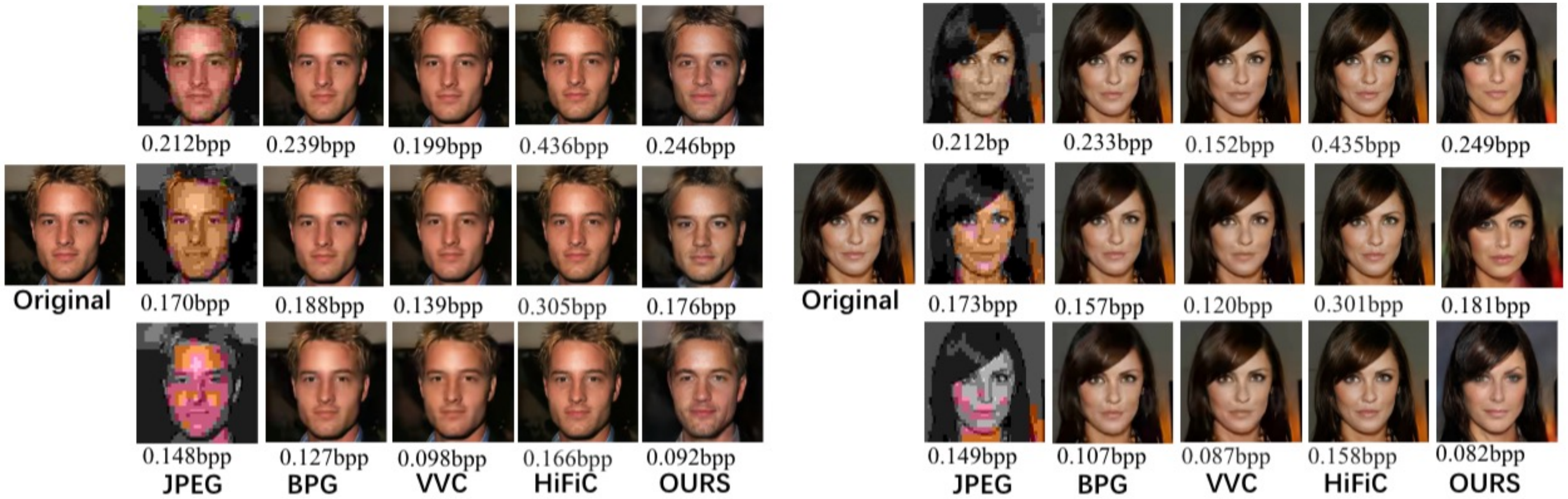
$$\mathcal{L}_D = \mathbb{E}_{x\sim p_X}[-\log(1 - D(x',y'))] + \mathbb{E}_{x\sim p_X}[-\log(D(x,y'))]$$

Analysis Loss:

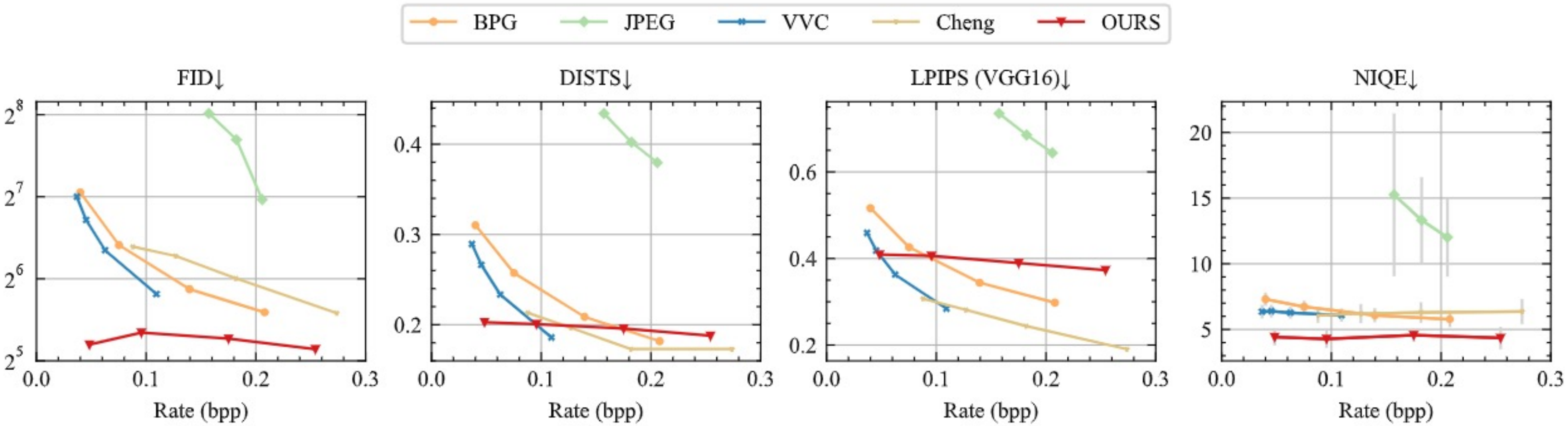
$$\mathcal{L}_{multi} = \lambda_{cls}l_{cls} + \lambda_{seg}l_{seg}$$

$$L = \underline{L_{distortion} + L_{adversarial}} + L_{rate} + L_{multitasks}$$

Experimental Results



Subjective Image Quality



Objective Image Quality

Experimental Results

Input Type	Codec	Rate (bpp)	Accu. (%)
Raw RGB Image	/	24*	90.06
Reconstructed RGB Image	BPG (qp=40)	0.075	88.55
	BPG (qp=37)	0.208	88.61
	JPEG (q=2)	0.157	87.10
	JPEG (q=3)	0.182	88.14
Compressed Latents	Ours (extreme)	0.048	86.74
	Ours (low)	0.096	87.74
	Ours (middle)	0.175	87.76
	Ours (high)	0.254	87.74

Face Attribute Prediction

Input Type	Codec	bpp	mIoU	FwIoU	Accu. (%)
RGB Images	/	24	0.692	0.876	93.29
Compressed Latents	Ours (extreme)	0.048	0.557	0.790	88.07
	Ours (low)	0.096	0.572	0.805	89.01
	Ours (middle)	0.175	0.580	0.814	89.57
	Ours (high)	0.254	0.598	0.822	90.07

Face Parsing

Experimental Results

Input Type	Codec	Rate (bpp)	Accu. (%)
Raw RGB Image	/	24*	90.06
Reconstructed RGB Image	BPG (qp=40)	0.075	88.55
	BPG (qp=37)	0.208	88.61
	JPEG (q=2)	0.157	87.10
	JPEG (q=3)	0.182	88.14
Compressed Latents	Ours (extreme)	0.048	86.74
	Ours (low)	0.096	87.74
	Ours (middle)	0.175	87.76
	Ours (high)	0.254	87.74

Face Attribute Prediction

Input Type	Codec	bpp	mIoU	FwIoU	Accu. (%)
RGB Images	/	24	0.692	0.876	93.29
Compressed Latents	Ours (extreme)	0.048	0.557	0.790	88.07
	Ours (low)	0.096	0.572	0.805	89.01
	Ours (middle)	0.175	0.580	0.814	89.57
	Ours (high)	0.254	0.598	0.822	90.07

Face Parsing

Input	Methods	Accu. (%)	#Param
RGB Images	Liu (2015b)	87	100M
	MOON (2016)	90.94	136M
	PSE (2018b)	91.23	62M
	AFFACT (2017)	91.67	26M
	DMTL (2017)	92.1	65M
Compressed Latents	Single Task	88.1	26.9M
	Multitasks	90.0	54.8M

Model Parameter Comparison

Analysis on Compressed Domain: A Multi-Task Learning Approach

Yuefeng Zhang[1], Chuanmin Jia[1], Jianhui Chang[1], Siwei Ma[1,2,3]
1. Peking University, 2. Peng Cheng Laboratory,
3. Information Technology R&D Innovation Center of Peking University

Data Compression Conference