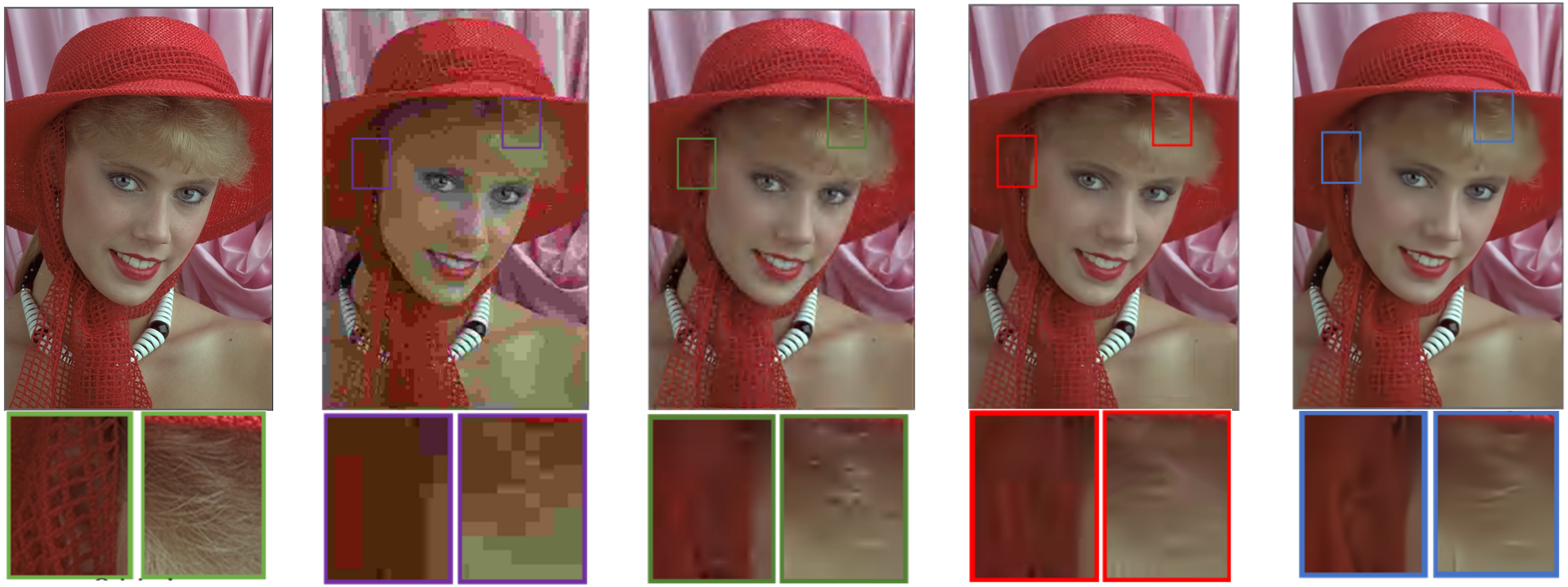
Fidelity-Controllable Extreme Image Compression with Generative Adversarial Networks

Shoma Iwai, Tomo Miyazaki, Yoshihiro Sugaya, Shinichiro Omachi
Tohoku University, Japan

Deep Image Compression

- Image compression is an important technique for efficient image storage and transmission.
- Recently, a lot of deep-learning based image compression methods have been studied.
- Some methods outperform conventional codecs such as JPEG, JPEG2000 and BPG.



	JPEG	JPEG2000	BPG	**Cheng et al. (2020)**
Orignal				
Bit rate	0.106bpp	0.109bpp	0.106bpp	**0.106bpp**
PSNR	24.26dB	28.59dB	30.03dB	**30.72dB**

Rate-Distortion Optimization

- Most deep image compression methods are trained to optimize the rate-distortion trade-off.

$$\mathcal{L} = R + \lambda D$$

Rate

Distortion
(e.g. MSE, MS-SSIM)

- However, especially at low bit rate, these methods suffer from blur.

Cheng et al. 2020
0.060bpp

original

GAN-based Method

- Some methods adopt GAN framework to reconstruct sharper images.
- However, GAN-based methods have two drawbacks.
  1. Training becomes unstable.
  2. Reconstructions often contain undesirable noise or artifact.

original

Agustsson et al. (2019)

Proposed Method

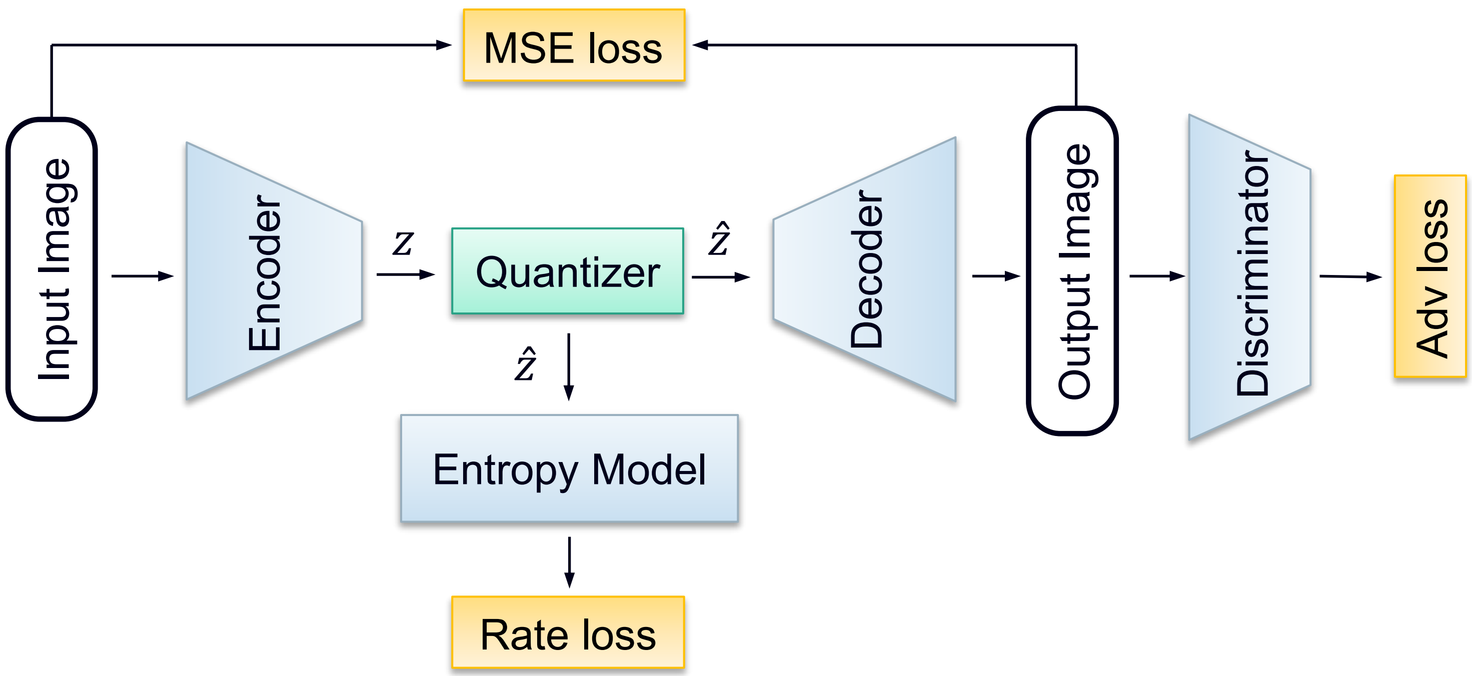
- We propose two strategies for these problems.

1. Two-Stage Training
   - Train the whole model without GAN.
   - Fine-tune only the decoder with GAN.

2. Network Interpolation
   - Merge two decoders (1^{st} and 2^{nd} stages) to reduce noise.

Our Compression Model

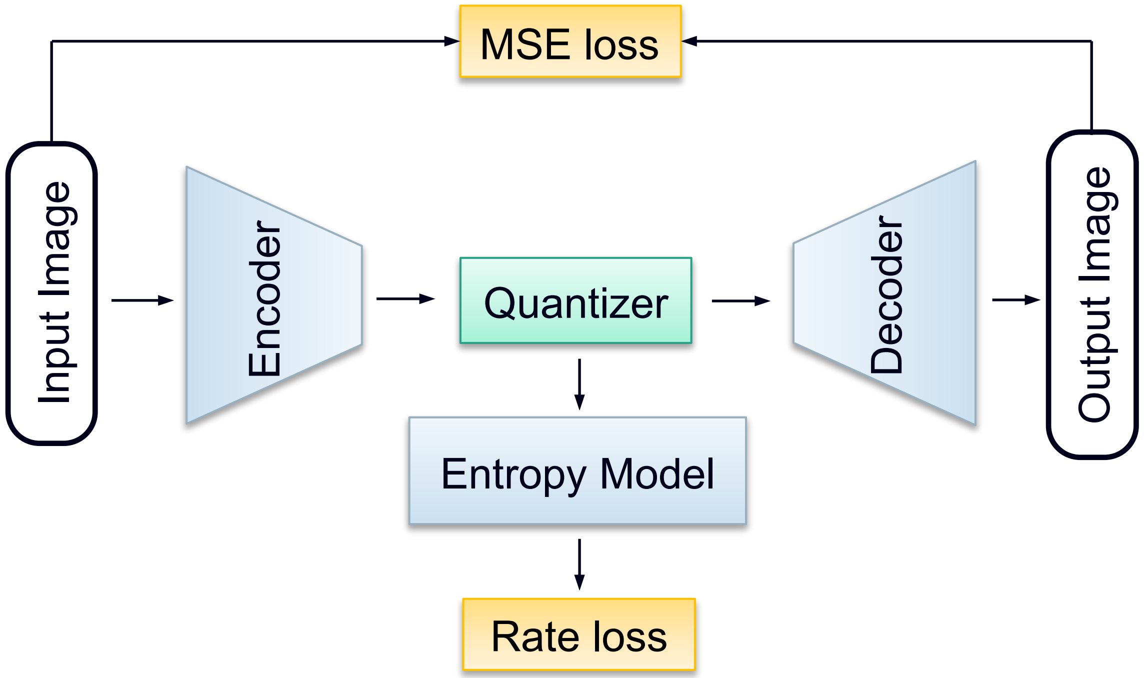
- **Encoder** transforms the input image into latent code z.
- **Quantizer** quantizes z into quantized code $\hat{z}$.
- **Decoder** reconstructs the image from $\hat{z}$.
- **Entropy model** estimates the bit rate of $\hat{z}$.
- **Discriminator** distinguishes the real image from the reconstruction.



Two Stage Training

1. Train all modules without GAN

2. Fine-tune only decoder with GAN

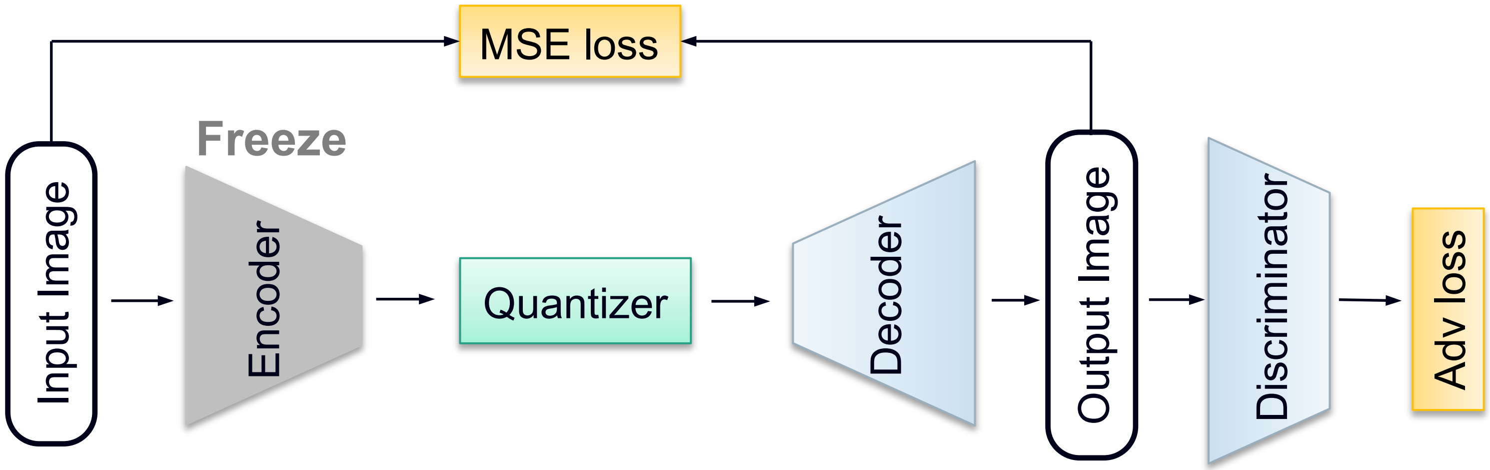


$$\mathcal{L}^{1st} = \mathcal{L}_{\text{Rate}} + \lambda \mathcal{L}_{\text{MSE}}$$

Two Stage Training

1. Train all modules without GAN

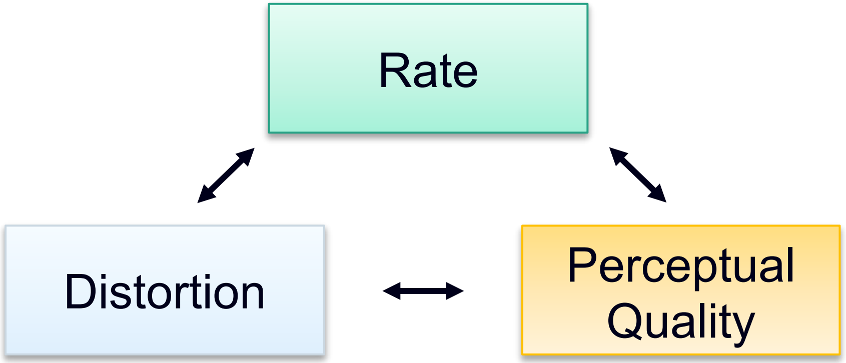
2. Fine-tune only decoder with GAN



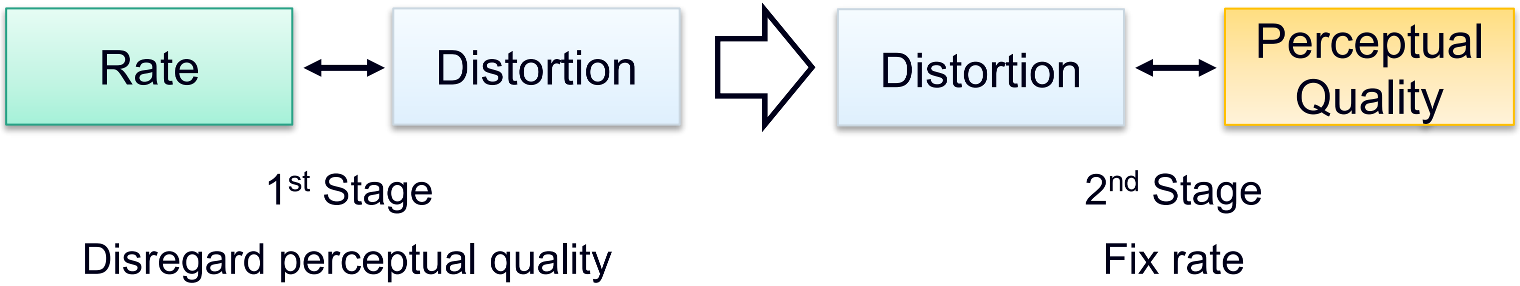
$$\mathcal{L}^{2nd} = \mathcal{L}_{\text{MSE}} + \lambda \mathcal{L}_{\text{Adv}}$$

Why Two Stage Training Work ?

- According to Blau et al. (2019), there is a triple trade-off between rate, distortion, and perceptual quality.



- The two-stage training relaxes optimization by splitting the triple trade-off.



Network Interpolation

- After training, we have two decoders:
  - 1st stage : High fidelity and Low perceptual quality
    low distortion but blurry
  - 2nd stage : Low fidelity and High perceptual quality
    sharp but contains noise

→ Merge two decoders to reconstruct visually more pleasing images

- Inspired by ESRGAN (Wang et al. 2018), we interpolate all the corresponding parameters of the two decoders.

$$\theta_{G'} = (1-\alpha)\theta_{G^1} + \alpha\theta_{G^2}$$

$\theta_{G'}$: Parameters of the new decoder

θ_{G^1}: Parameters of the decoder in the 1st stage

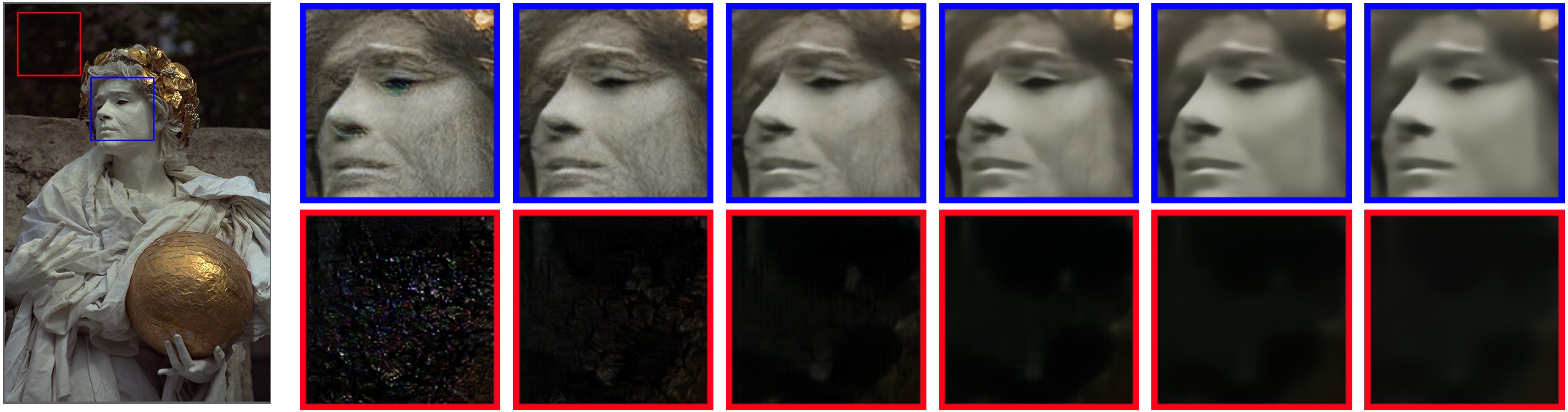
θ_{G^2}: Parameters of the decoder in the 2nd stage

$\alpha \in [0, 1]$: interpolation parameter

Fidelity Control by Network Interpolation

- We can control the trade-off between distortion and perceptual quality by adjusting α without re-training the model.

	2nd stage					1st stage
	$\alpha = 1.0$	$\underline{\alpha = 0.8}$	$\alpha = 0.6$	$\alpha = 0.4$	$\alpha = 0.2$	$\alpha = 0$



PSNR [dB]	24.00	24.83	25.23	25.69	26.19	26.36
Bit rate [bpp]	0.031	0.031	0.031	0.031	0.031	0.031

Comparison with Existing Methods

- BPG and Cheng et al. (state-of-the-art PSNR-oriented model) suffer from blur.
- Agustsson et al. contain artifacts.
- Our reconstruction looks natural.

original

Ours 0.031bpp

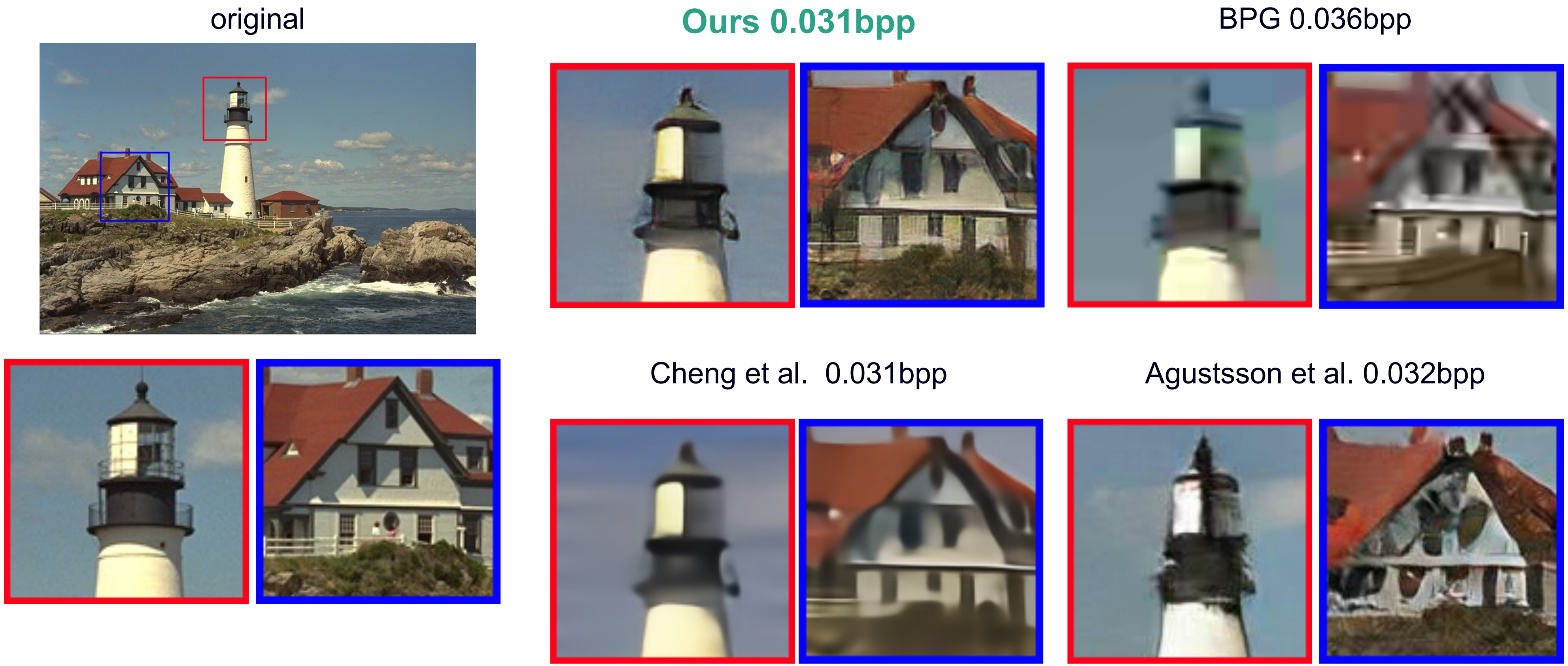
BPG 0.036bpp

Cheng et al. 0.031bpp

Agustsson et al. 0.032bpp

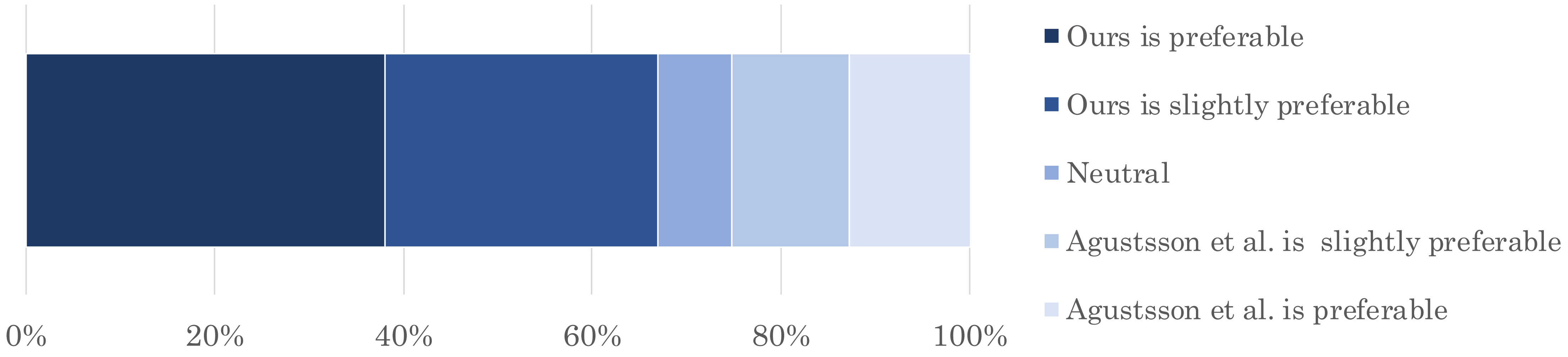
Comparison with Existing Methods

- BPG and Cheng et al. (state-of-the-art PSNR-oriented model) suffer from blur.
- Agustsson et al. contain artifacts.
- Our reconstruction looks natural.



User Study

- We performed a user study to compare our method with Agustsson et al. (2019).
- We asked 19 users to evaluate which reconstruction is preferable.
- More than 60% of the answers are ‘Ours is preferable’ or ‘Ours is slightly preferable’.



Conclusion

- We proposed a GAN-based extreme image compression method.
- We adopt the two-stage training and the network interpolation to tackle the two problems of GAN-based methods.
- Our reconstructions are perceptually high quality.
- Our user study shows the proposed method outperforms state-of-the-art GAN-based method, Agustsson et al.

Thank you for your attention !