Fidelity-Controllable Extreme Image Compression with Generative Adversarial Networks

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

Original Image

Ours
0.031bpp

Agustsson et al. (2019)
0.033bpp

BPG
0.032bpp
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.
Most deep image compression methods are trained to optimize the rate-distortion trade-off.

\[ \mathcal{L} = R + \lambda D \]

Rate\hspace{1cm}Distortion
\hspace{1cm}(e.g. MSE, MS-SSIM)

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

Cheng et al. 2020
0.060bpp

original
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.
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 (1st and 2nd 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

\[ \mathcal{L}^{1st} = \mathcal{L}_{\text{Rate}} + \lambda \mathcal{L}_{\text{MSE}} \]
1. Train all modules without GAN

2. Fine-tune only decoder with GAN

\[ \mathcal{L}^{2nd} = \mathcal{L}_{MSE} + \lambda \mathcal{L}_{Adv} \]
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.

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.
After training, we have two decoders:

- 1\textsuperscript{st} stage: \textbf{High fidelity} and \textbf{Low perceptual quality}
  low distortion but blurry

- 2\textsuperscript{nd} stage: \textbf{Low fidelity} and \textbf{High perceptual quality}
  sharp but contains noise

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

\[
\theta_G' = (1 - \alpha)\theta_G^1 + \alpha \theta_G^2
\]

- Parameters of the new decoder
- Parameters of the decoder in the 1\textsuperscript{st} stage
- Parameters of the decoder in the 2\textsuperscript{nd} stage

\(\alpha \in [0, 1] : \text{interpolation parameter}\)
We can control the trade-off between distortion and perceptual quality by adjusting $\alpha$ without re-training the model.

Fidelity Control by Network Interpolation

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>2nd stage PSNR [dB]</th>
<th>2nd stage Bit rate [bpp]</th>
<th>1st stage PSNR [dB]</th>
<th>1st stage Bit rate [bpp]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1.0$</td>
<td>24.00</td>
<td>0.031</td>
<td>26.36</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha = 0.8$</td>
<td>24.83</td>
<td>0.031</td>
<td>26.19</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha = 0.6$</td>
<td>25.23</td>
<td>0.031</td>
<td>25.69</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha = 0.4$</td>
<td>25.69</td>
<td>0.031</td>
<td>25.23</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha = 0.2$</td>
<td>25.69</td>
<td>0.031</td>
<td>25.23</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha = 0$</td>
<td>26.00</td>
<td>0.031</td>
<td>25.69</td>
<td>0.031</td>
</tr>
</tbody>
</table>
BPG and Cheng et al. (state-of-the-art PSNR-oriented model) suffer from blur.

Agustsson et al. contain artifacts.

Our reconstruction looks natural.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Ours 0.031bpp</th>
<th>BPG 0.036bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1" alt="Original Image" /></td>
<td><img src="image2" alt="Ours 0.031bpp" /></td>
<td><img src="image3" alt="BPG 0.036bpp" /></td>
</tr>
<tr>
<td></td>
<td><img src="image4" alt="Cheng et al. 0.031bpp" /></td>
<td><img src="image5" alt="Agustsson et al. 0.032bpp" /></td>
<td></td>
</tr>
</tbody>
</table>
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.
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’.
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.