Controlling Rate, Distortion, and Realism: Towards a Single Comprehensive Neural Image Compression Model

Zixuan Chen
Mentor: Till Aczel
Seminar in Deep Neural Networks
26.03.2024, ETH Zurich
Introduction: Image Compression

Usage: image storage and transmission

Three main indicators: Rate, distortion, and realism
Guess and discuss

Which two are of the same bit rate?
Which one maintains the most details?

From Iwai et al. [*thisp]
Bit-rate: BPP (bits per pixel)

From Iwai et al. ["thisp"]
Bit-rate

From Iwai et al. [*thisp]
Measurements of Distortion

$$MSE(f, g) := \frac{1}{|V|} \int_V |f(x) - g(x)|^2 \, dx \quad (\text{Mean Squared Error})$$

$$PSNR(f, g) := 10 \log_{10} \left( \frac{m^2}{MSE(f, g)} \right) \quad (\text{Peak Signal to Noise Ratio})$$

V: a rectangular region of the image
f, g: images
m: the maximum possible pixel value of the image

From Becker Axel [*measure]
Distortion

lower distortion = closer to the original image

Realism ≠ distortion

From Wang, et al. [*sameMSE*]
Measurements of Realism

1. FID
2. LPIPS
   - Both use deep convolutional networks.
Realism

low FID

~ high realism

From Fan et al. [*realism]
High Realism itself makes no sense
We want **low** bit rate, **low** distortion, and **high** realism!

However, these three indicators **cannot** be achieved simultaneously!
Rate-distortion-realism Tradeoff
(Curves on the blackboard)

- Fix rate R
- Fix distortion D
- Fix FID’s upper bound P

From Yochai Blau and Tomer Michaeli [*tradeoff]
Tradeoff: takeaway messages

- At low bit rates, the tradeoff becomes stronger.
- To optimize one metric, the other two need to be sacrificed.

From Yochai Blau and Tomer Michaeli [*tradeoff]
A Classic Compression Pipeline:

Single-rate, no realism control
Single-rate v.s. Variable-rate
Loss function
Previous Work

• Learning based:
  - Generative:
    - GAN-based: Multi-realism, HiFiC, PQMIM
    - Diffusion-based: HFD, DIRAC
  - Non-Generative: ELIC, Charm, IVR, Hyperprior
• Non-Learning based: VTM, JPEG

“Green”: able to adjust Distortion-realism tradeoff in one model
GAN Based Training
Motivation: how to adjust the balance between rate, distortion, and realism within a single model?
This Paper: Pipeline
(See Blackboard)

beta: realism weight (higher beta means higher realism and higher distortion, vice versa)
This Paper: Loss function

\[ \mathcal{L}_{1st} = \lambda_R^{(q)} R(\hat{y}_q) + \lambda_d d(x, \hat{x}_q) + \mathcal{L}_P(x, \hat{x}_q) \]

bit rate,    MSE,    LPIPS

\[ \mathcal{L}_{2nd} = \lambda_R^{(q)} R(\hat{y}_q) + \lambda_d d(x, \hat{x}_q) + \beta(\lambda_P \mathcal{L}_P(x, \hat{x}_q) + \lambda_{adv} \mathcal{L}_{HRRGAN}^G) \]

bit rate,    MSE,    LPIPS,    adversarial loss
To control the rate:
Insert Interpolation Channel Attention Layers

This page till the end: from Iwai et al. [*thisp]
**Discriminator - RaGAN**

Relativistic Average GAN

\[
p_r(x_r, x_f) = \sigma(D(x_r) - \mathbb{E}_{x_f}[D(x_f)])
\]

\[
p_f(x_r, x_f) = \sigma(D(x_f) - \mathbb{E}_{x_r}[D(x_r)])
\]

\[
\mathcal{L}_{RaGAN}^G = -\log p_f(x_r, x_f) - \log(1 - p_r(x_r, x_f))
\]

\[
\mathcal{L}_{RaGAN}^D = -\log p_r(x_r, x_f) - \log(1 - p_f(x_r, x_f))
\]
Discriminator - RGAN
Relativistic GAN

\[ \mathcal{L}_{RGAN}^G = - \log \sigma(D(x_f) - D(x_r)) \]

\[ \mathcal{L}_{RGAN}^D = - \log \sigma(D(x_r) - D(x_f)). \]
Discriminator - HRRGAN

Higher Rate Relativistic GAN

To avoid over-penalty on realism

\[ \mathcal{L}^G_{\text{HRRGAN}} = - \log \sigma(D(\hat{x}_q) - \text{sg}(D(\hat{x}_{q+1}))) \]
\[ \mathcal{L}^D_{\text{HRRGAN}} = - \log \sigma(D(x) - D(\hat{x}_q)), \]

sg: stop gradient operation
Independent vs Shared Discriminator

(a) Independent

(b) Shared

: (1) Conv layer applied for all quality levels
Hybrid Discriminator

backbone: extract and encode features

head: produce prediction

(c) Hybrid-head
(d) Hybrid-backbone

(2) Conv layer applied for a specific quality level
Experimental Results: Compare with Generative Models

<table>
<thead>
<tr>
<th>Original</th>
<th>(bpp, PSNR)</th>
<th>HiFiC (single-rate)</th>
<th>Multi-Realism (variable rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.311bpp, 20.6dB</td>
<td>0.401bpp, 23.2dB</td>
</tr>
</tbody>
</table>

**Ours:** Low-rate, Low-distortion  
$q = 0, \beta = 0$  
0.308bpp, 23.1dB

**Ours:** Low-rate, High-realism  
$q = 0, \beta = 3.84$  
0.308bpp, 22.7dB

**Ours:** High-rate, Low-distortion  
$q = 4, \beta = 0$  
2.24bpp, 34.5dB

better texture!

Comparable realism with lower bit rate!
Quantitative Evaluation

PSNR↑ (distortion) [CLIC2020]

FID↓ (realism) [CLIC2020]

<table>
<thead>
<tr>
<th>Variable-rate</th>
<th>Single-rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours β=0.0 (low-distortion)</td>
<td>Ours w/o MR trained with fixed β = 2.56</td>
</tr>
<tr>
<td>Ours β=3.84 (high-realism)</td>
<td>DIRAC-1 (arxiv'23)</td>
</tr>
<tr>
<td>DIRAC-100 (arxiv'23)</td>
<td>VTM</td>
</tr>
<tr>
<td>Multi-Realism β=0.0 (CVPR'23)</td>
<td>Multi-Realism β=2.56 (CVPR'23)</td>
</tr>
<tr>
<td>HiFiC (NeurIPS'20)</td>
<td>HFD (arxiv'23)</td>
</tr>
<tr>
<td>PQ-MIM (TMLR'23)</td>
<td></td>
</tr>
</tbody>
</table>

bpp(bit rate)
Quantitative Evaluation

- Perform fine rate-tuning

- high realism model: surpassed DIRAC on both indicators
Results of Different Discriminator designs

- Hybrid discriminators outperformed shared discriminators in FID
- Quality-level specific layers are beneficial
Effect of HRRGAN

- Average calculation harms performance

Trained with fixed beta = 2.56
RGAN: Relativistic GAN
SGAN: Standard GAN
RaGAN: Relativistic Average GAN
HRRGAN: Higher Rate Relativistic GAN
Limitation

- Control the rate and realism uniformly
  cannot perform precise (e.g. pixel-level) control
References


