Adaptive Image Compression Method Using GAN based Semantic-Perceptual Residual Compensation

Present. Wang Ruojing
Outline

◆ Introduction
  ❖ Background
  ❖ Related work
  ❖ Motivation

◆ Proposed method
  ❖ GAN based Dense Auto-encoder (GDAE)
    ● DAE structure
    ● GAN
  ❖ Semantic-perceptual Residual Compensation (SPRC)
    ● Semantic-perceptual Method: Grad-CAM
    ● Residual Compensation Block

◆ Experiments
◆ Conclusion & Future work
◆ Reference
I. Introduction
Introduction-background

◆ Why Image compression:
  ◆ (1) limited storage space  (2) narrowband communication channels
  ◆ Traditional image compression algorithm:
  ◆ JPEG/JPEG2000/BPG(lossy), PNG/GIF/FLIF(lossless).....

Kodak Image datasets: 768*512

24bit RGB ground truth:
24*768*512 = 9437184bit
  = 1152KB
  = 1.25M

◆ Why we need deep learning methods?
  ◆ (1) compressed quality  (2) compressed rate -> rate-distortion optimization
  ◆ The merits of DL compression network?
  (1) Extract structure information from the data. -> Reduce data redundancy efficiently
  (2) Predict the semantic information from the human vision->content adaptive
Introduction-background

- JPEG and JPEG2000
  - Distortion problem (blurred edge, noise) of traditional algorithms

JPEG  0.125[bit/pixel]  JPEG2000  0.125[bit/pixel]
Introduction-related work

- Typical Auto-encoder based compression structure
  - Distortion problem is still remained even if use deep learning methods

Fully connected layers based Auto-encoder

Fully convolutional layers based Auto-encoder

Input Data

Quantization & storage

Reconstruct Data
### Introduction-related work

<table>
<thead>
<tr>
<th>work</th>
<th>Publication</th>
<th>method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lossy compressive AE [1]</td>
<td>2017 ICLR</td>
<td>CNN model</td>
</tr>
<tr>
<td>Full resolution with RNN[2]</td>
<td>2016 CVPR</td>
<td>RNN model</td>
</tr>
</tbody>
</table>

![Diagram of image compression process]

- **CNN Encoder**
- **Quantization**
- **Inverse Quantization**
- **CNN Decoder**
- **Entropy Coding**
- **Code Prediction**
- **Reconstruction Error**
- **Rate-distortion Optimization**
Introduction-related work

Examples of the Problem

① Deep CNN are more likely to lose details due to it’s several downsampling operations.

② Reconstructed images often do not conform to human vision.

③ The images generated by the neural network are not naturally distributed and prone to distortion in some local areas.

Abnormal distribution generated by the neural network

Humans tend to pay attention to the foreground rather than the background, but neural networks often fail to achieve it.

In this example we want the reconstructed image can retain more details about the parrot.
Introduction-related work

- Heat map (the distribution of the detected region)
  - comparison of various methods of detecting objects in the image

The Grad-CAM method is the improved CAM.
- Such methods can be easily combined with neural networks to mimic human attention and predict regional importance.
Motivation

- **Motivation:**
  - Introduce the ideal of generative adversarial neural network (GAN)

MSE can promote the pixel-wised numerical approximation between reconstructed image and original image. **But it does not consider the structural similarity and statistical distribution similarity of the whole image.**

Through the adversarial learning between the generator and discriminator GAN makes the reconstructed image statistically fit to the distribution of the real image.
Motivation

**Motivation:**

- Implement GAN based dense Auto-encoder, the discriminator can push the decoder to generate the image statistically closer real distribution
- **Semantic-perceptual method** will compensate the unstability of common CNN for the lost details from the perspective of human vision

---

Our whole method combines existing methods and several well-designed novel strategies to achieve comparable compression performance with the state of the art methods.
II. Proposed methods
GDAE

- For compression Auto-encoder
  - encoder
  - quantizer
  - decoder

- Dense connection
  - Idea: Inspired by U-net(left), similar densely residual connections are used on the compressed network.
Densely connected auto-encoder details

The skip connection design can extract much more feature information.
Compressed Auto-Encoder Based on GAN

- GAN structure: discriminator and generator

- The discriminator will give a score (0~1) comparing the input image and the reconstructed image. The decoder will learn from the score.

- The discriminator also plays the role of generator.

- Range 0~1, 0 means the reconstructed image is totally different from the ground truth, 1 is contrast.
GDAE

◆ Loss function

❖ General loss function that GAN used[7]
❖ SSIM loss[8]
❖ MSE loss[9]

General loss function in GAN:
\[ l_{GAN} = E_{x\sim q(x)} \log D(x) + E_{z\sim p(z)} \left[ \log \left( 1 - D(G(z)) \right) \right] \]

SSIM loss:
\[ l_{SSIM} = -I(x, \hat{x}) \cdot C(x, \hat{x}) \cdot S(x, \hat{x}) \]

Illuminance, Contrast, Structural information

\[ I(x, \hat{x}) = \frac{2\mu_x \mu_{\hat{x}} + C_1}{\mu_x^2 + \mu_{\hat{x}}^2 + C_1} \]
\[ C(x, \hat{x}) = \frac{2\sigma_x \sigma_{\hat{x}} + C_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2} \]
\[ S(x, \hat{x}) = \frac{\sigma_{x\hat{x}} + C_3}{\sigma_x \sigma_{\hat{x}} + C_3} \]

MSE loss:
\[ l_{MSE} = \frac{1}{N} \| x - \hat{x} \| \]

x means input image, x-hat means the reconstructed image.

Total loss:
\[ Loss = \lambda_1 L_{MSE} + \lambda_2 L_{GAN} + \lambda_3 L_{SSIM} \]

◆ Total loss: \( \lambda_1 : \lambda_2 : \lambda_3 = 10:5:2 \)
GDAE

GDAE demo result

- Choose one image from the testing dataset Kodak PhotoCD, bpp around 0.25.

The instability of the GAN network makes the generated reconstructed image have color distortion.

For improve the subjective quality of compressed image. We design a Semantic-perceptual residual compensation block (SPRC).
Heat map generated by Grad-CAM

The class of parrot will be activate.

Grad-CAM method is the improved CAM, the shortcoming is that CAMs are trained to maximize posterior probability for only one class, they tend to only highlight a single most prominent object.
SPRC

◆ SPRC is for subjective compensation task
  ▶ a sub-work separated from the GDAE

◆ VGG-16 use the pre-training parameters of 1000 image classification tasks on ImageNet, and then fine-tuning in combination with GAN.
More details on the residual image

- a sub-work separated from the GDAE

BPG can be flexible choose 1-51 compression level, so we use **linear mapping**, according to the score range, adapt the compression quality.
FLIF is the latest lossless codec, is to reduce the memory data after quantization process.
III. Experiments
Experiments

◆ Experiment datasets
  ❖ Training Dataset: ImageNet
  ❖ Testing Dataset: Kodak PhotoCD

◆ Enviroment
  ❖ GPU: Dual NVIDIA RTX2080Ti,
  ❖ CPU: Intel core-i7 8700K,
  ❖ Memory: DDR4 3200MHz

  The preprocessing steps are as follows:
  ❖ (1) Perform random scaling of 0.5 to 2.0 times and random cropping of the area.
  ❖ (2) Since an image of the same size should be input in a Batch in the GPU, we will uniformly collect the image to a size of 256 × 256.
Experiment result

- **GDAE: GAN+DAE+Soft-quantization**
  - `img`: input image; `rec`: reconstructed image.

![Experiment results](image-url)
Experiment result

- **GDAE+SPRC**
  - The hat shows that the color distortion problem be solved by SPRC.
  - The layers of the local cat's fur, residual compensation really worked.
Experiment result

Rate-distortion results

- Dataset: Kodak PhotoCD

- The results of **PSNR in 0.4-1.5 bpp** have achieved marginal improvement among the state of art traditional and deep learning methods. The results of **MS-SSIM in 0.5-1.6 bpp** have exceeded them all.
Experiment result

Ablation analysis

- Table 1 choose the bpp around 0.5, calculate the PSNR and MS-SSIM.
- Table 2 AE means Auto-encoder, DC means dense connection.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR</th>
<th>MS-SSIM(dB)</th>
<th>RT(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG [2]</td>
<td>29.3</td>
<td>13.7</td>
<td>290</td>
</tr>
<tr>
<td>BPG [5]</td>
<td>32.9</td>
<td>15.2</td>
<td>750</td>
</tr>
<tr>
<td>Lee’s [18]</td>
<td>33.2</td>
<td>19.3</td>
<td>6200</td>
</tr>
<tr>
<td>Our’s(GDAE)</td>
<td>33.1</td>
<td>18.6</td>
<td>607</td>
</tr>
<tr>
<td><strong>Our’s(GDAE+SPRC)</strong></td>
<td><strong>33.8</strong></td>
<td><strong>20.1</strong></td>
<td><strong>912</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AE</th>
<th>DC</th>
<th>GAN</th>
<th>SPRC</th>
<th>PSNR</th>
<th>MS-SSIM(dB)</th>
<th>RT(ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>27.37</td>
<td>16.91</td>
<td>426</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>29.24</td>
<td>17.07</td>
<td>532</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>32.65</td>
<td>17.96</td>
<td>532</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>32.47</td>
<td>18.66</td>
<td>863</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>30.77</td>
<td>18.21</td>
<td>426</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>30.92</td>
<td>18.28</td>
<td>752</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>31.27</td>
<td>18.94</td>
<td>752</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>33.77</td>
<td>19.24</td>
<td>863</td>
</tr>
</tbody>
</table>

Lee's work is also using the deep learning method, our running time is better than his.
IV. Conclusion & future work
Conclusion & future work

- Our work achieves:
  - firstly develop a GAN based dense autoencoder to make full use of the feature information extracted from the input image (GDAE).
  - Next, add a semantic-perceptual residual compensation block to GDAE architecture, lead to an improving compression performance (SPRC).

- Drawbacks
  - It's time-consuming when training the networks, 5000 epochs for a week.
  - Grad-CAM can only activate the most prominent class in an image.
Conclusion & future work

Future work

- More semantic-perceptual (objective detection) methods
  - MS-ROI and Aug-ROI

- Calculate the time that every separated part has used, including encoder, quantizer, decoder.

- Further optimize the quantization process to further reduce the distortion.


Q & A