WASEDA University
Graduate School of Information, Production and Systems

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

PNG(lossless)： 544KB
JPEG(lossy)： 107KB

- **Why we need deep learning methods？**
  - (1) **compressed quality** (2) **compressed rate** ->**rate-distortion optimization**
  - The merits of DL compression network？

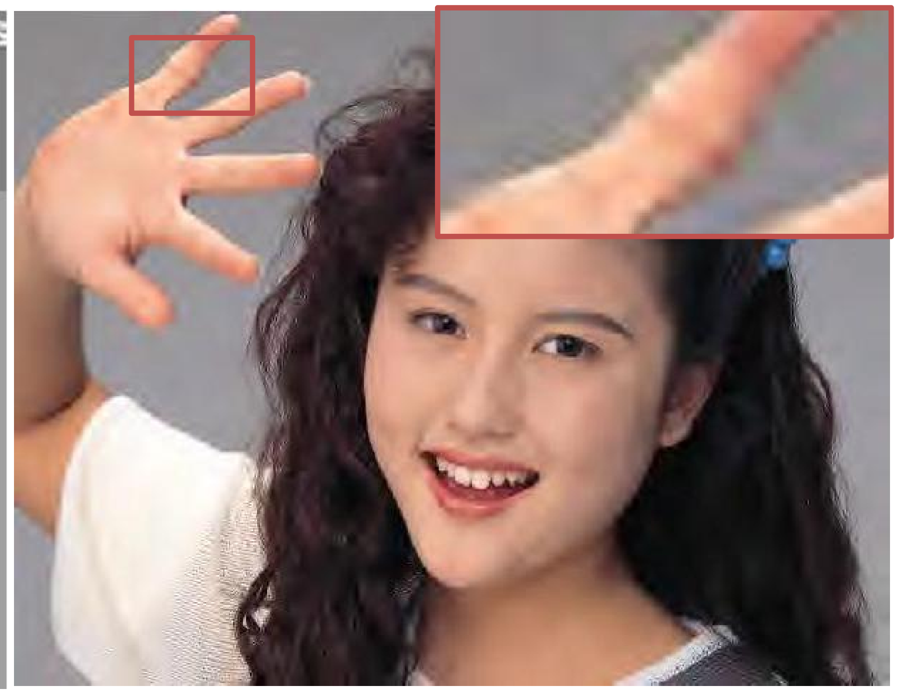
(1) Extract structure information from the data. -> **Reduce data redundancy efficiently**

(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]

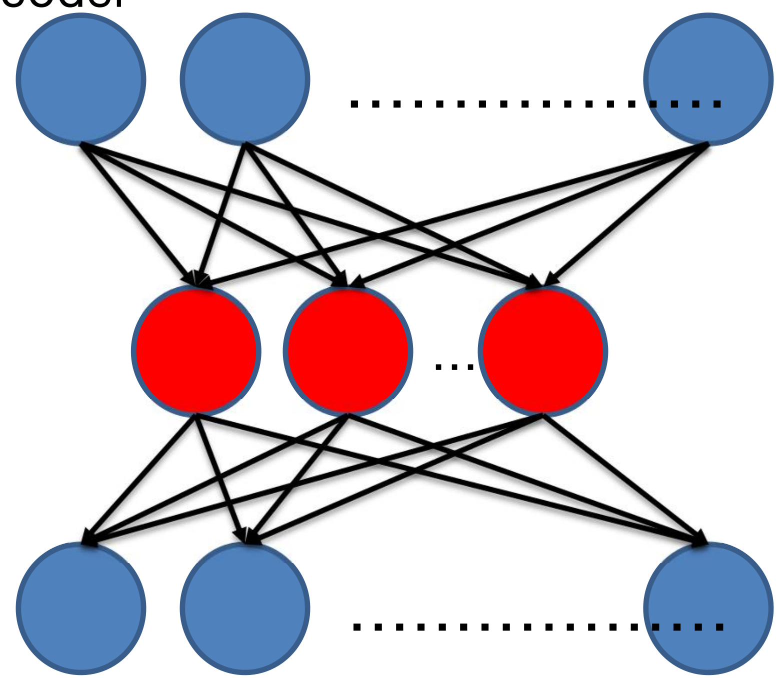


JPEG2000 0.125[bit/pixel]

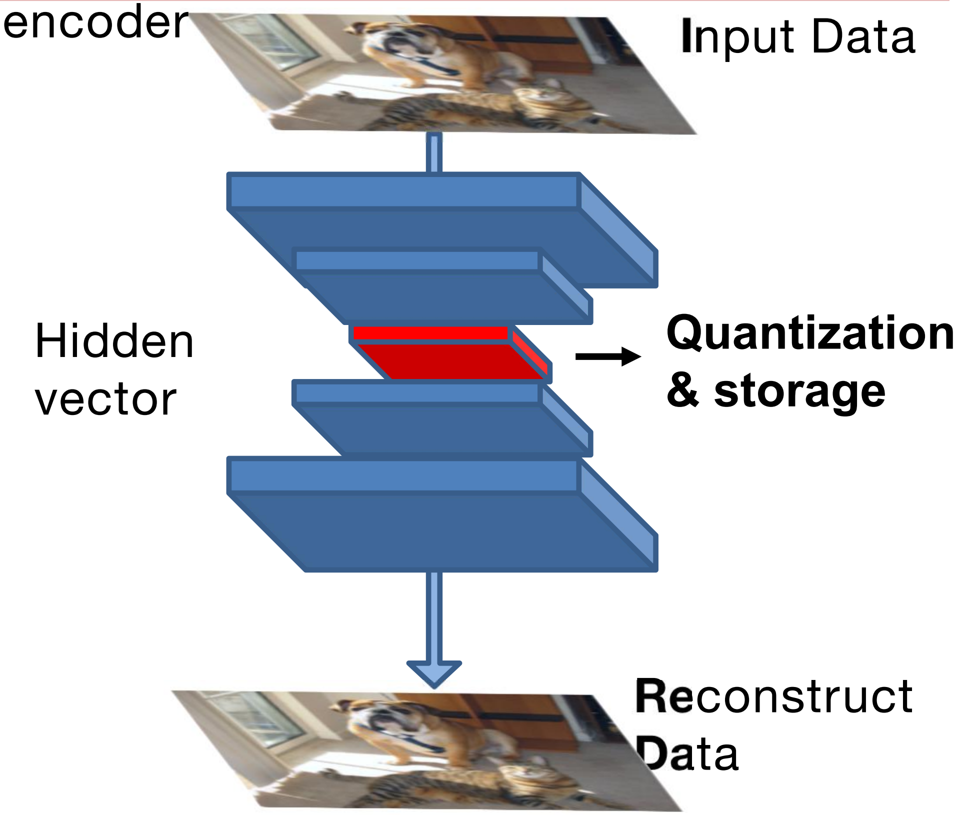
Introduction-related work

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

Fully connected layers based Auto-encoder

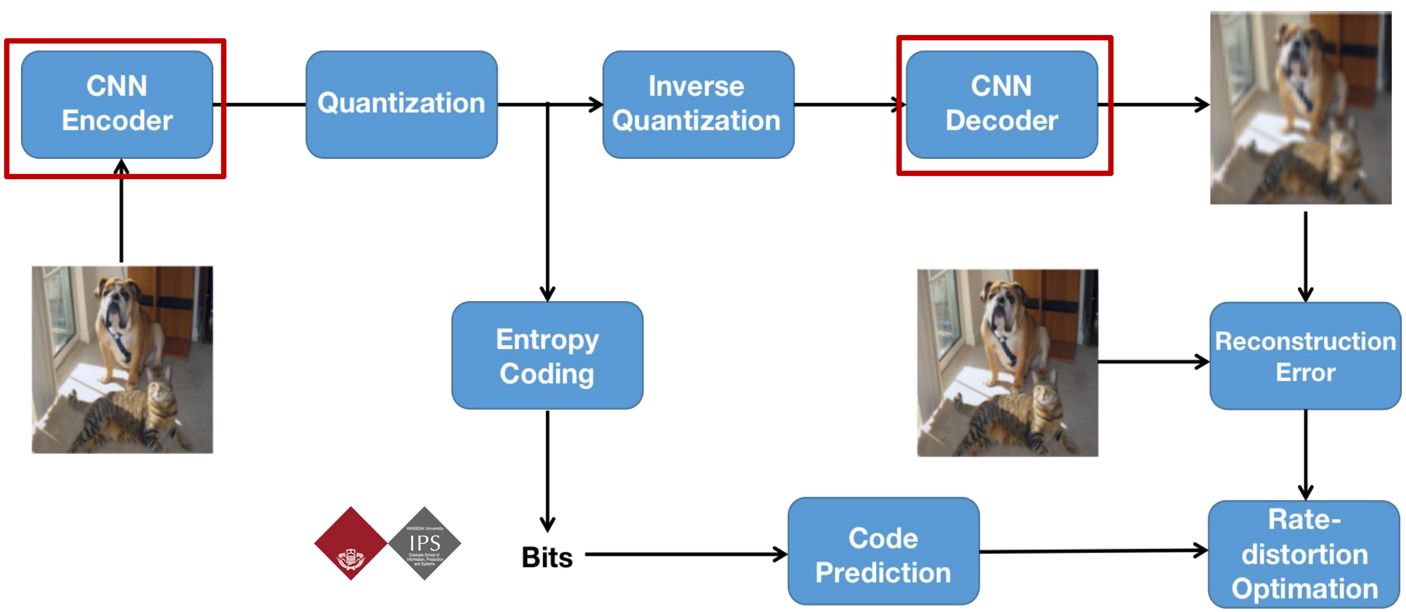


Fully convolutional layers based Auto-encoder



Introduction-related work

work	Publication	method
Lossy compressive AE [1]	2017 ICLR	CNN model
Full resolution with RNN[2]	2016 CVPR	RNN model
Smantic-perceptual net[3]	2017 DCC	CNN model and MS-ROI for adaptive JEPG codec
GAN for compression [4]	2018 ICCV	Graph neural network method
Deep semantic segmantation GAN[5]	2019 NSERC	GAN model and residual compensation.



Introduction-related work

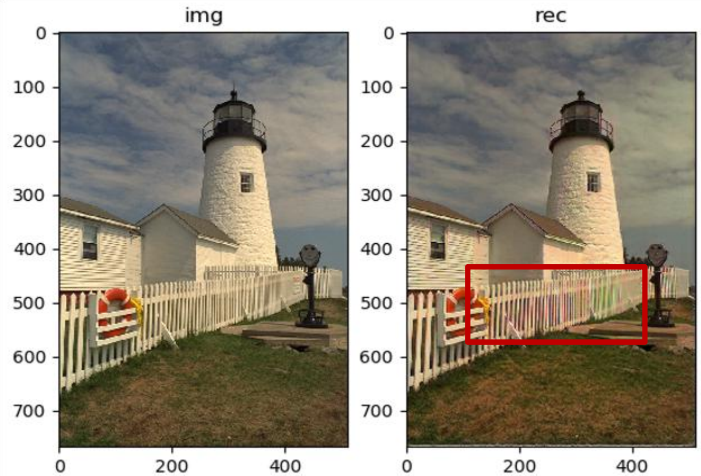
◆ Examples of the Problem

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 dis
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

A. Original

B. Human Fixation

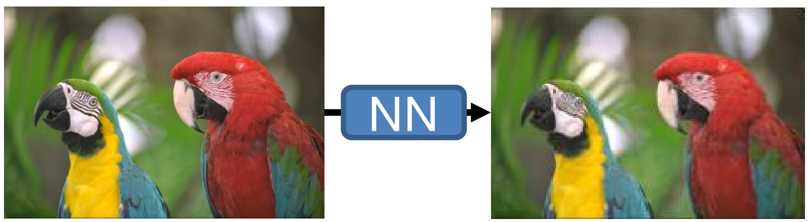
C. CAM

- 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:

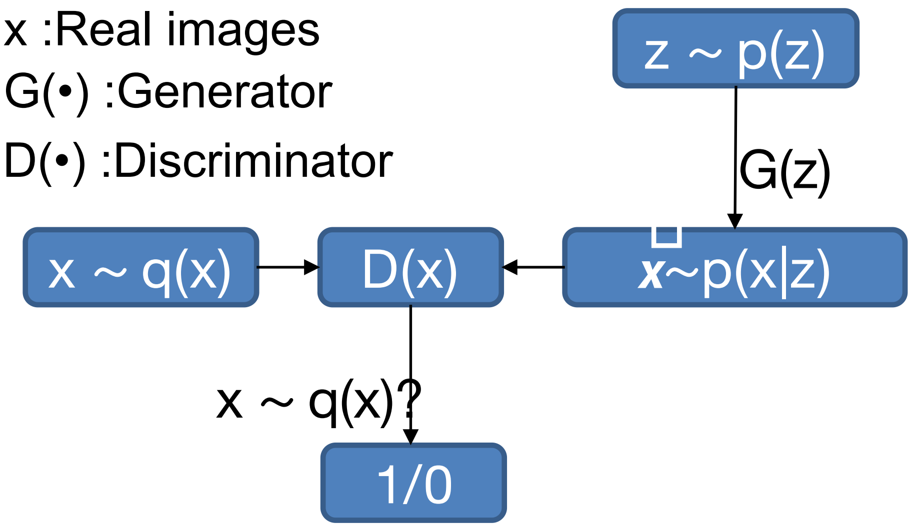
❖ Introduce the ideal of generative adversarial neural network (GAN)



Loss: Mean square error (MSE)

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**

$\hat{\boldsymbol{x}}$: Reconstructed Image
z : Hidden-vector
x :Real images
G(•) :Generator
D(•) :Discriminator

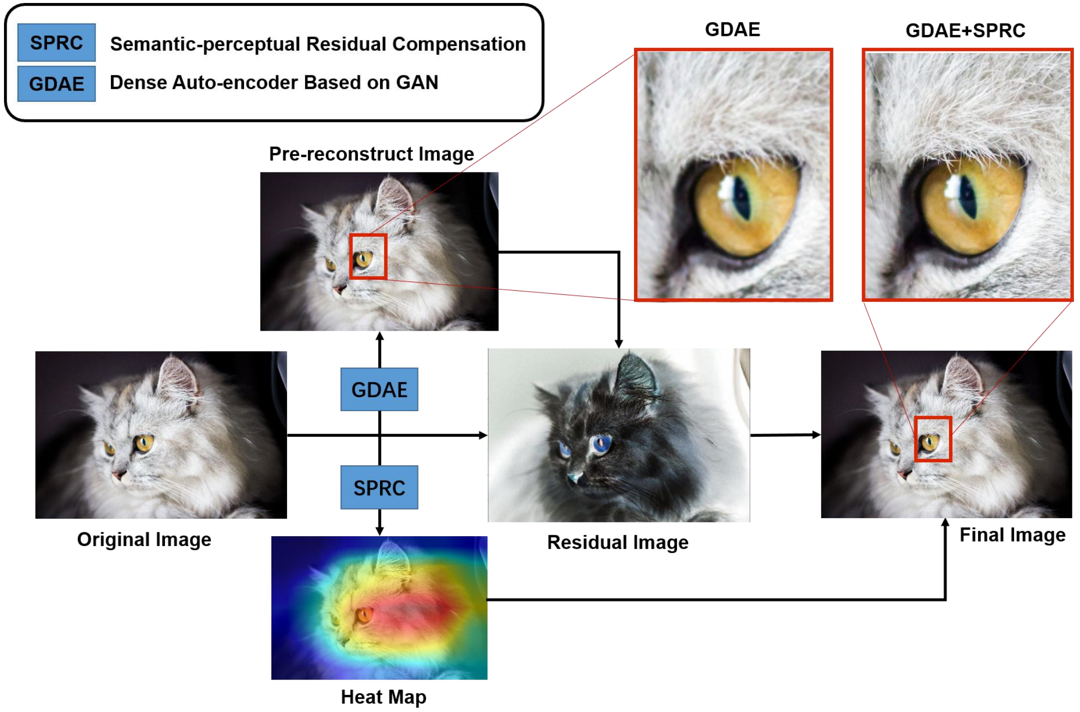


Through the adversarial learning between the generator and discriminator GAN makes the reconstructed image statistically fit to the distribution of the real image.

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

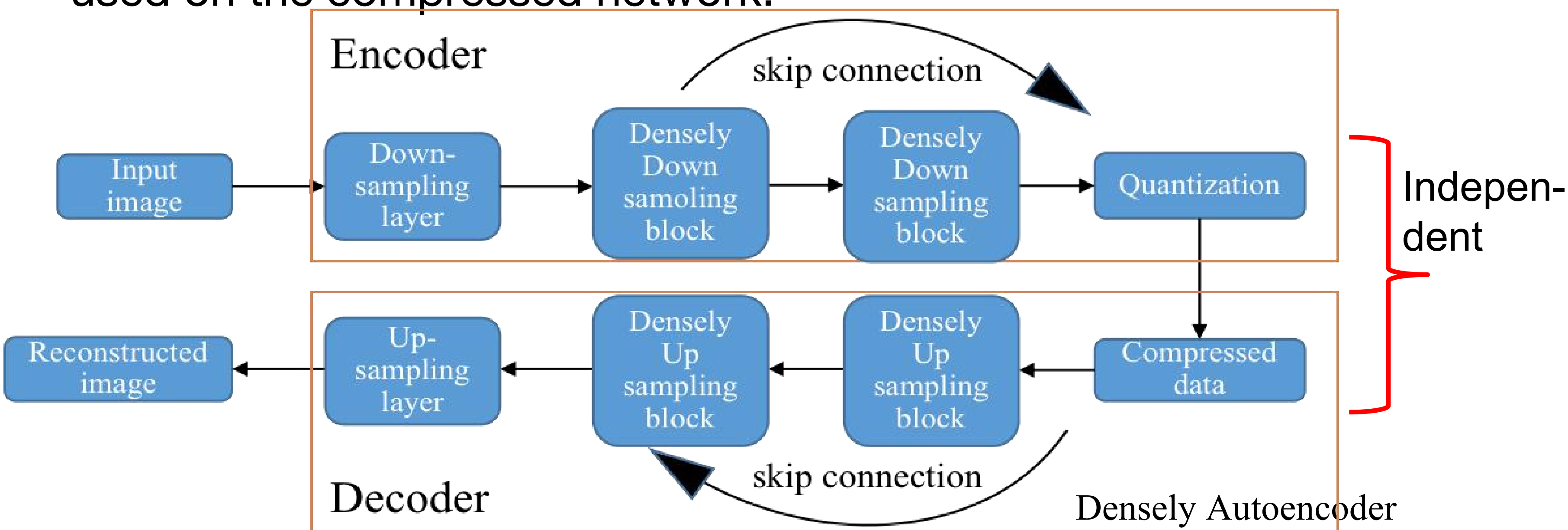
Our whole method combines existing methods and several well-designed novel strategies to achieve comparable compression performance with the state of the art methods.



II. Proposed methods

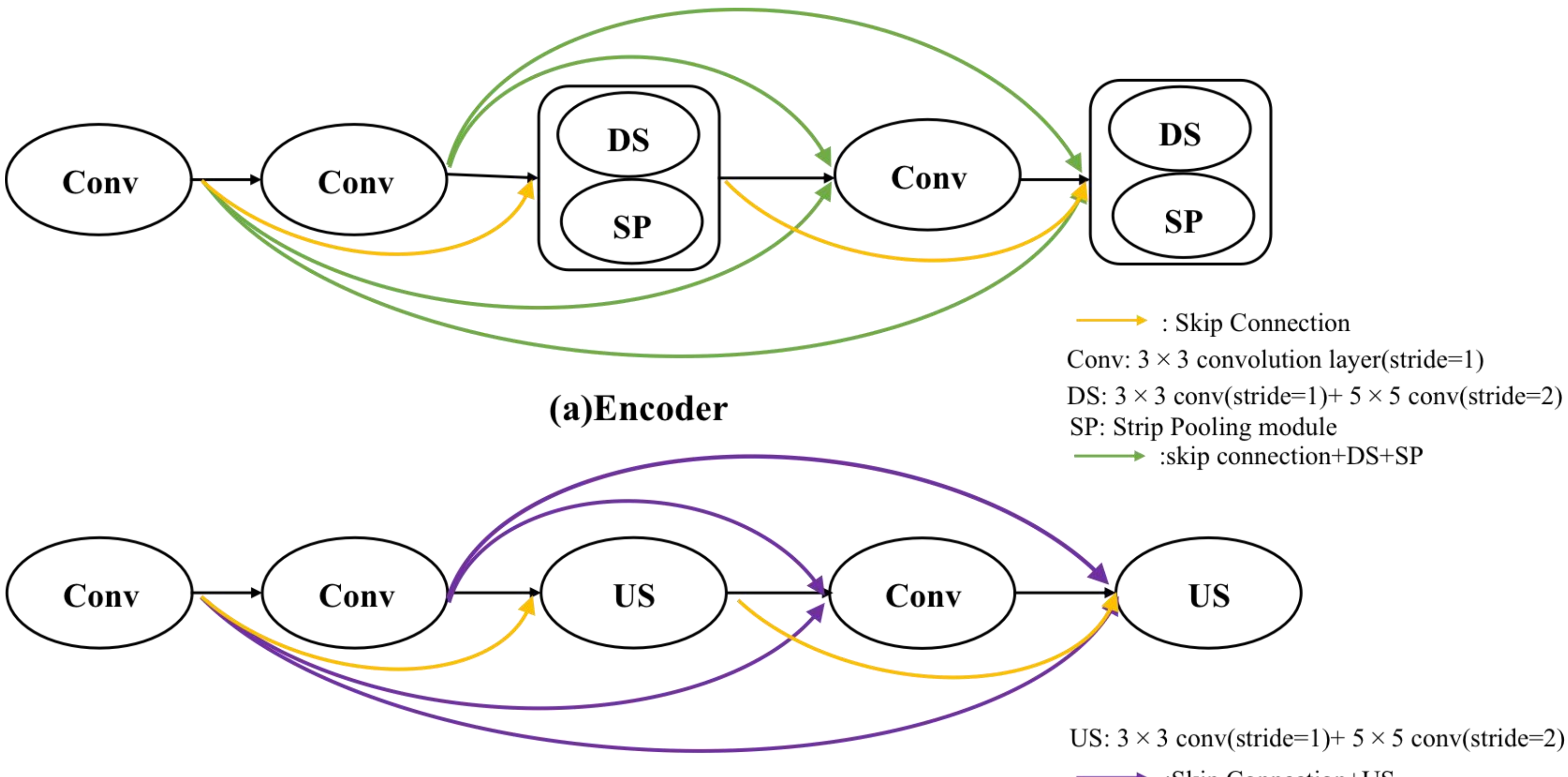
GDAE

- 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.



GDAE

- Densely connected auto-encoder details
  - The skip connection design can extract much more feature information



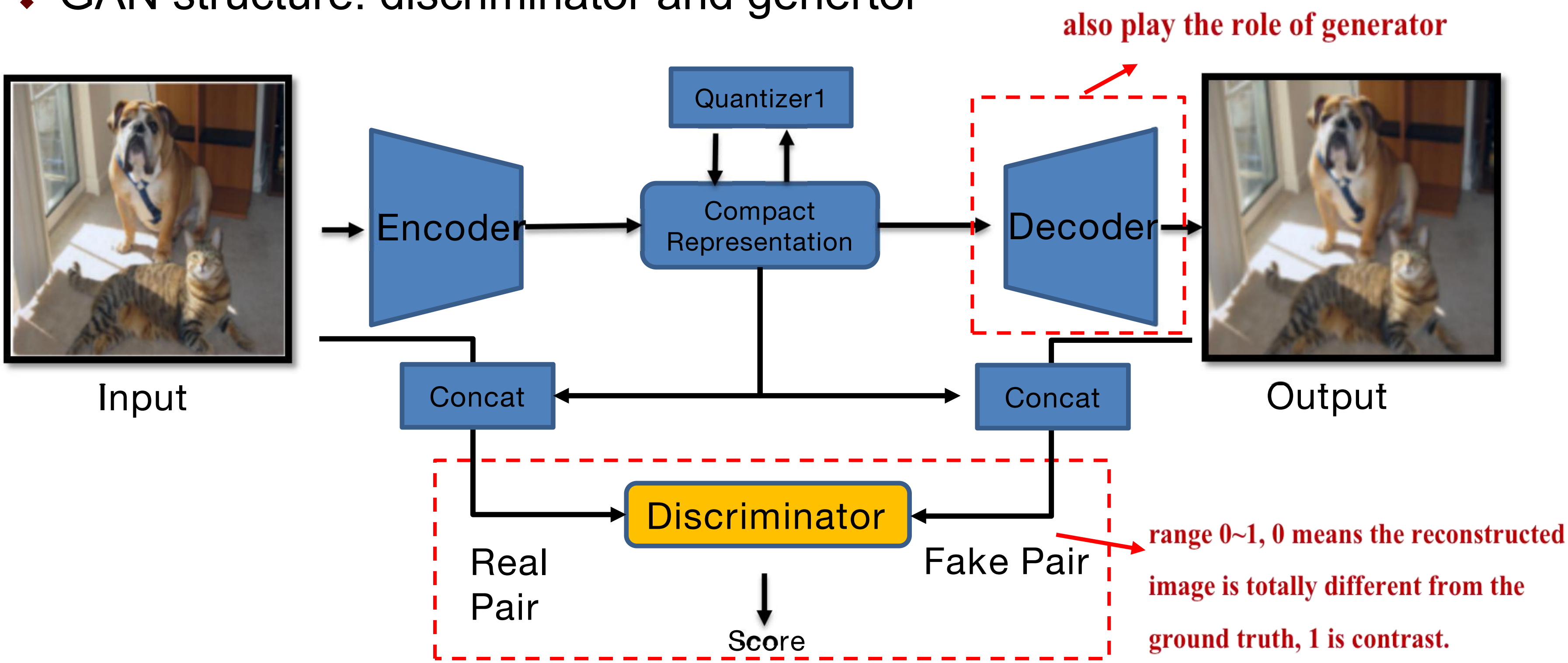
(a)Encoder

(b)Decoder

GDAE

◆ Compressed Auto-Encoder Based on GAN

❖ GAN structure: discriminator and genertor



❖ The discriminator will give a score(0~1) compare to the input image and the reconstructed image, the decoder will learn from the score.

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, Structurual information

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

MSE loss: $l_{\mathrm{MSE}} = \frac{1}{N}\|x - \hat{x}\|$

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 datasat Kodak PhotoCD, bpp around 0.25.

Ground Truth DAE, **PSNR=28.14** GDAE, **PSNR=29.68**

- 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 resitual compensation block(SPRC).

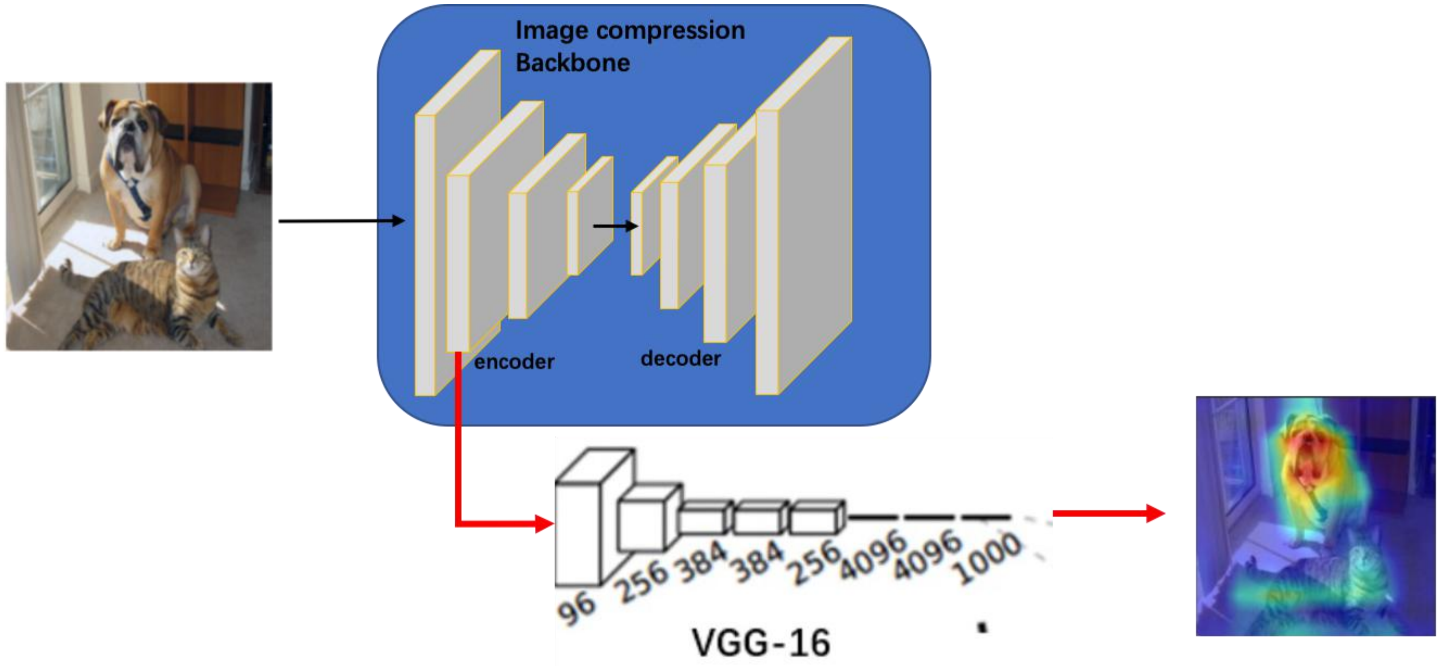
SPRC

- 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

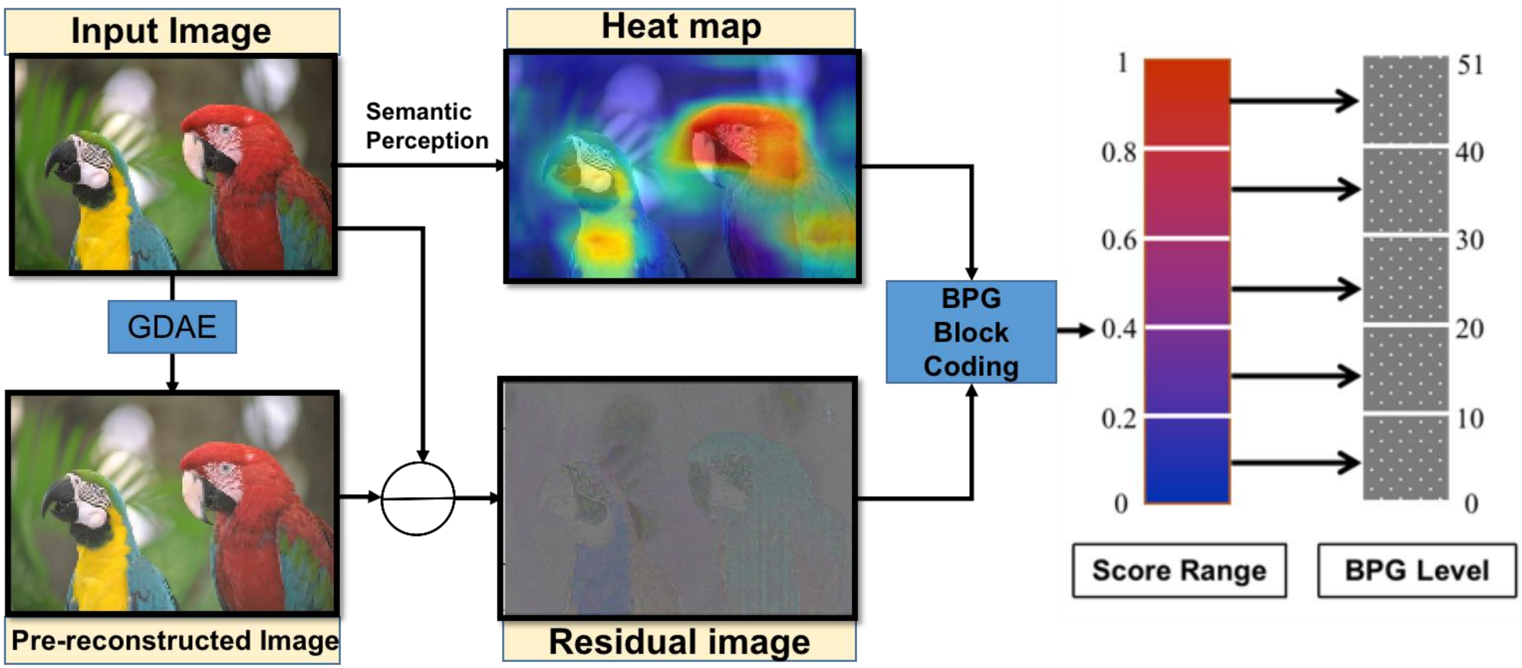
- SPRC is for subjective compensation task
  - a sub-work separated from the GDAE



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

SPRC

- More details on the residual image
  - a sub-work separated from the GDAE

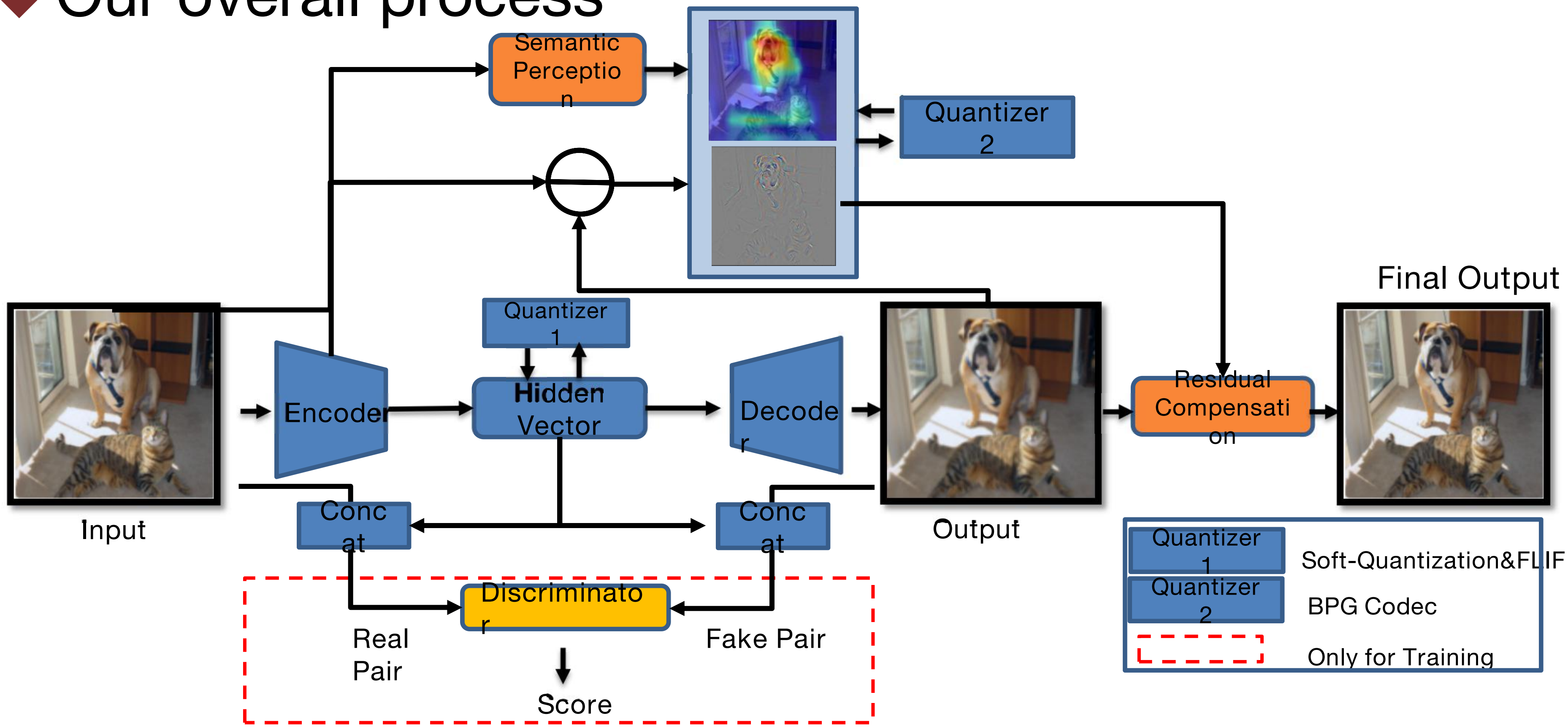


(a) SPRC Block (b) BPG Details

  - BPG can be flexible choose 1-51 compression level, so we use linear mapping, accoring to the score range, adapt the compression quality.

GDAE+SPRC

- Our overall process



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

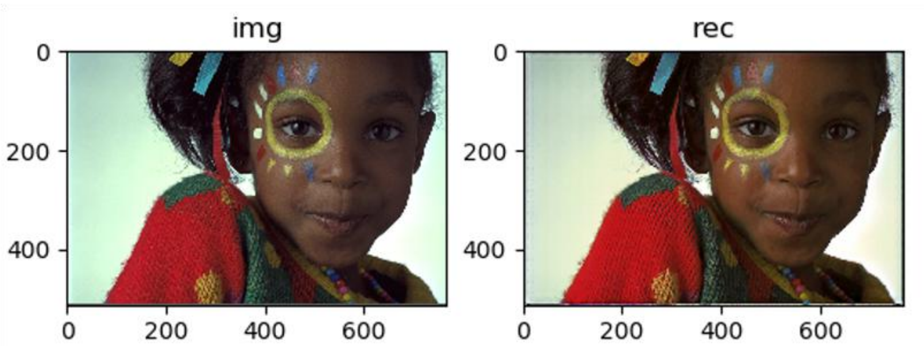
Kodak PhotoCD

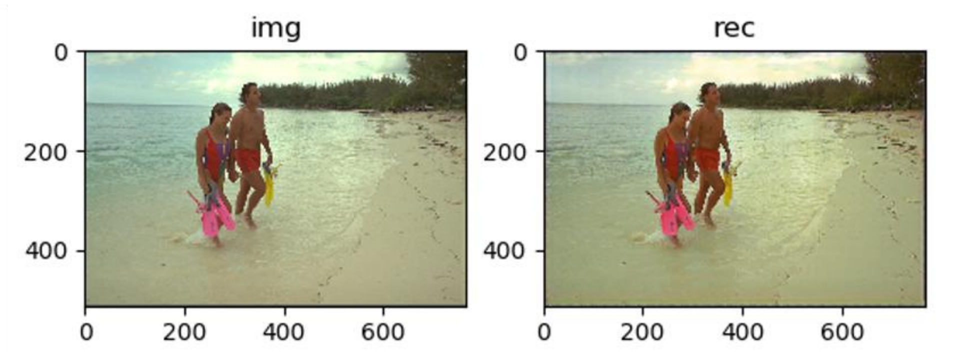
- ◆ Enviroment
  - GPU: Dual NVIDIA RTX2080Ti,
  - CPU: Intel core-i7 8700K,
  - Memory: DDR4 3200MHz

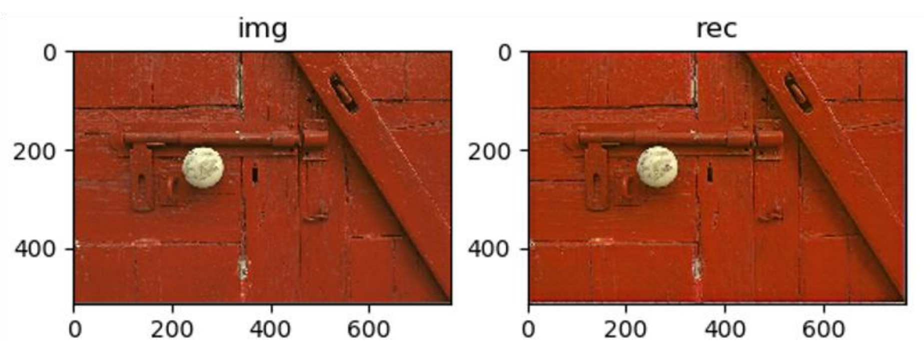
- The preprocessing steps are as follows:
- (1) Perform **random scaling of 0.5 to 2.0 times** and **random cropping** of the area.
- (2) Since an image of the same size should be input in a Batch in the GPU, we will uniformly collect the image to a **size of 256 × 256**.

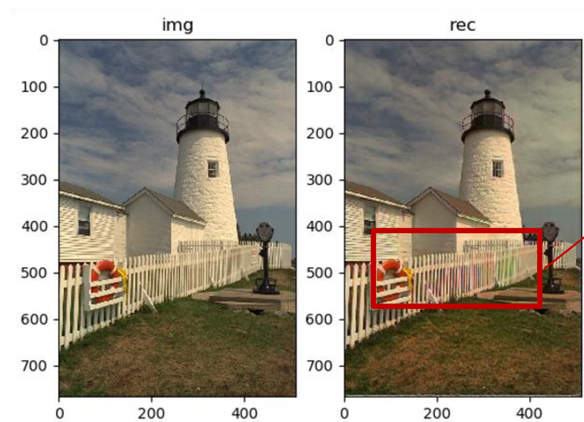
Experiment result

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





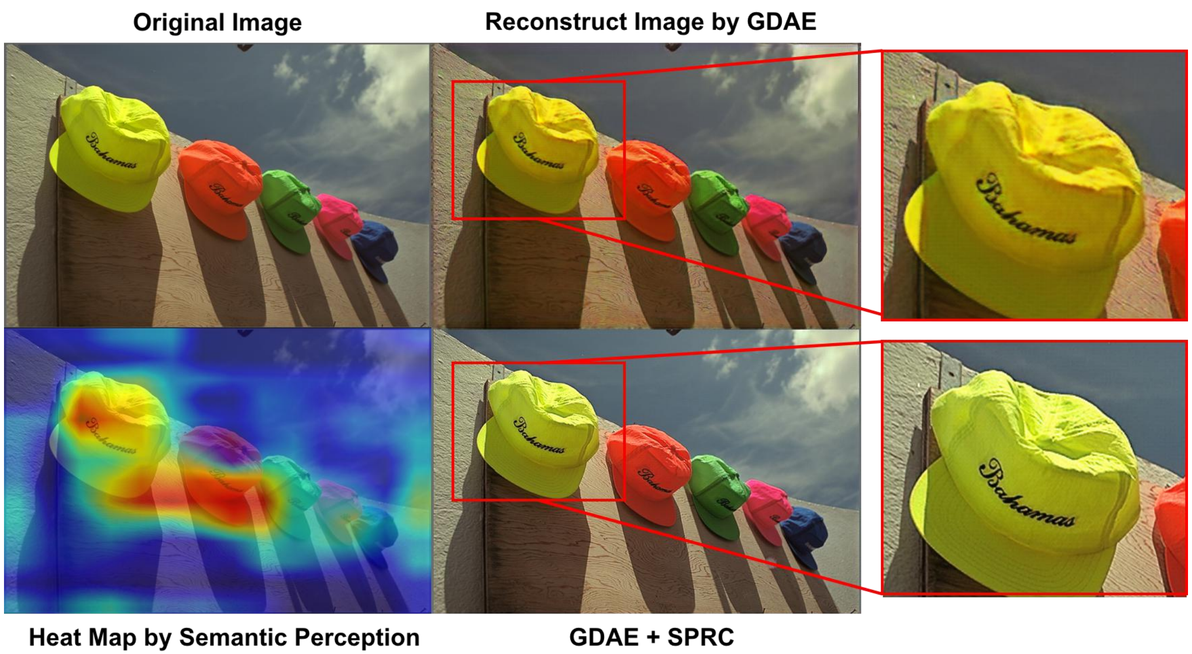




Color distortion

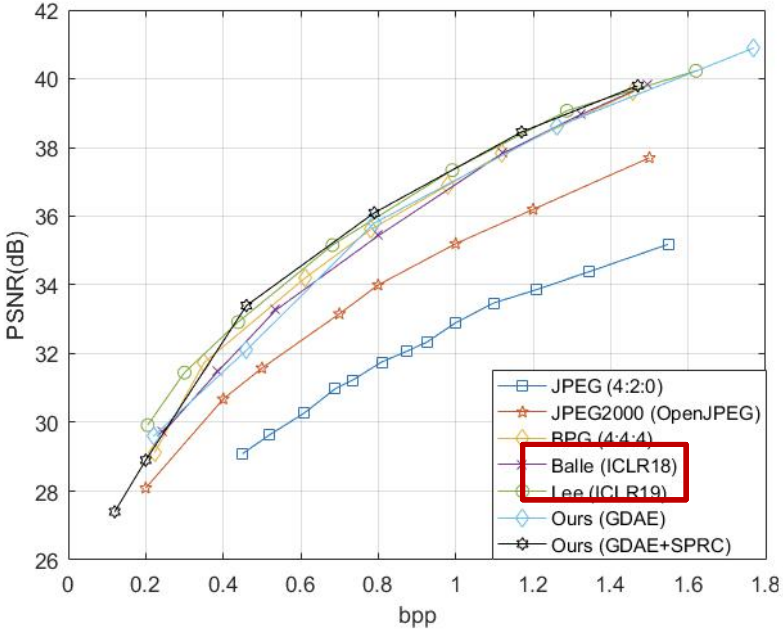
Experiment result

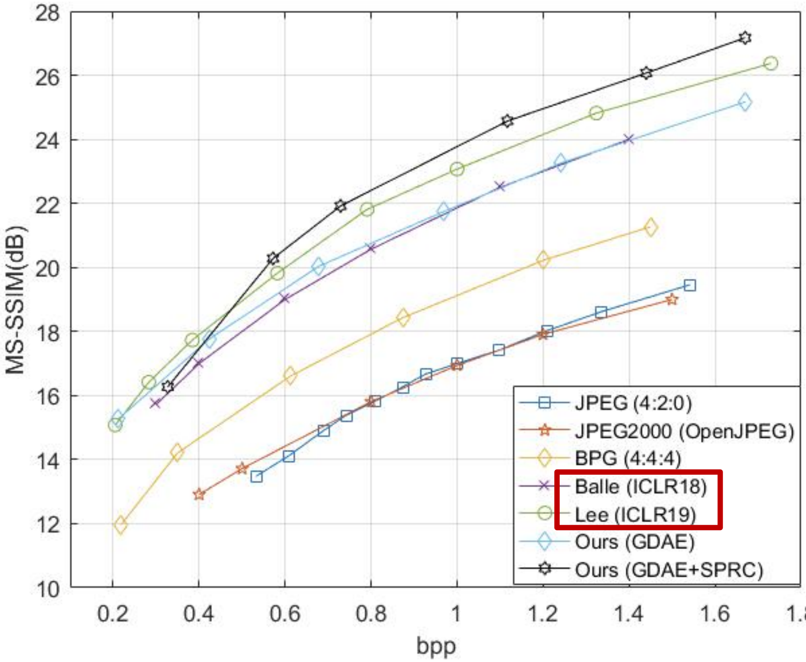
- GDAE+SPRC
  - The hat shows that the color distortion problem be solved by SPRC.
  - The layers of the local cat's fur, residual compensation really worked.



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 1

Method	PSNR	MS-SSIM(dB)	RT(ms)
JPEG [2]	29.3	13.7	**290**
JPEG2000 [3]	31.7	13.9	590
BPG [5]	32.9	15.2	750
Lee's [18]	33.2	19.3	6200
Our's(GDAE)	33.1	18.6	607
Our's(GDAE+SPRC)	**33.8**	**20.1**	912

Table 2

AE	DC	GAN	SPRC	PSNR	MS-SSIM(dB)	RT(ms)
✓				27.37	16.91	426
✓	✓			29.24	17.07	532
✓	✓	✓		32.65	17.96	532
✓	✓		✓	32.47	18.66	863
✓		✓		30.77	18.21	426
✓			✓	30.92	18.28	752
✓		✓	✓	31.27	18.94	752
✓	✓	✓	✓	33.77	19.24	863

- 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-comsuming** when training the networks, 5000 epochs for a week.
  - Grad-CAM can only activate the most prominant class in an image.

Conclusion & future work

Future work

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

Grad-CAM MS-ROI 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.

Reference

- [1] Liu J, Lu G, Hu Z, et al. “A Unified End-to-End Framework for Efficient Deep Image Compression[J]”. arXiv preprint arXiv:2002.03370, 2020.
- [2] He, Kaiming, et al. "Deep residual learning for image recognition." arXiv preprint arXiv:1512.03385 (2015).
- [3] J. Ball´e, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” in 6th International Conference on Learning Representations, ICLR, 2018.
- [4] J. Lee, S. Cho, and S.-K. Beack, “Context-adaptive entropy model for end-to-end optimized image compression,” arXiv preprint arXiv:1809.10452, 2018.
- [5] P. Aaditya, et al. “Semantic Perceptual Image Compression using Deep Convolution Networks”, Data Compression Conference (DCC),2017.
- [6] Hou, Qibin, et al. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [7] A. Eirikur, et al. “Soft-to-hard vector quantization for end-to-end learning compressible representations”. Advances in Neural Information Processing System, 2017.
- [8] G. Ian, et al. “Generative adversarial nets”,Advances in neural information processing systems, 2014.
- [9] S. Shibani, et al. “Generative compression”, Picture Coding Symposium (PCS), 2018.
- [10] O. Rippel, et al. “Real-time adaptive image compression”, Proceedings of the 34th International Conference on Machine Learning, {ICML} 2017.

Q & A