ROI Perceptual Metric

Yi Ma, Yongqi Zhai, Jiayu Yang, Chunhui Yang, Ronggang Wang
Shenzhen Graduate School, Peking University, China
{mayi,yangchunhui,jiayuyang}@pku.edu.cn
{zhaiyongqi}@stu.pku.edu.cn {rgwang}@pkusz.edu.cn

Abstract

With the continuous development of image compression technology, people gradually focus on improving the visual perception quality of reconstructed images. However, people have different standards for different types of texture evaluation, which makes it unreasonable to use spatially unified standards to evaluate the reconstructed image. For this reason, we propose a ROI Perceptual metrics. For the saliency region, we require the reconstructed features to be more accurate. For the background region, we want texture features to be more realistic and not to reject generative details. In addition, We use a multi-layer perceptron network to fuse different types of metrics to save the process of manually adjusting parameters.

1. Method

We will generate the basic metrics and ROI metrics of the two images to be evaluated, and then input them into the multilayer perceptron (MLP) to obtain the final prediction.

1.1. ROI Mask

First, we generate saliency regions through a saliency detection network[1]. Please note that we will fix this network as a strong supervision for training the compression network, rather than co-training with it. Next, the saliency map is smoothed by convolution(the filter size is 50, the values are all set to 1) to generate a 2D ROI mask RM_{2D}.

$$RM_{2D} = S(\sigma(N(x))) \tag{1}$$

where N denotes the saliency detection network, S denotes the smooth convolution, and σ refers to the sigmoid function.

1.2. Metrics

The metrics we use can be divided into two types: basic metrics and ROI metrics. Basic metrics include PSNR, MS-SSIM, LPIPS and discriminator loss:

$$L_{basic} = \{L_{MSE}, L_{MS-SSIM}, L_{LPIPS}, L_{D-LOSS}\} \tag{2}$$

Due to their simplicity, PSNR and MS-SSIM are widely used in the field of image compression. LPIPS calculates the distance of the image in the feature space of a deep neural network. We participated in the image compression track and got a well-trained image compression model based on GAN. Discriminator loss comes from that model. ROI metrics only calculate the differences in the ROI regions:

$$L_{ROI} = mean(L_{basic} \otimes RM_{2D}) \tag{3}$$

1.3. Multilayer Perceptron

As shown in the Figure 1, Our network input is a 16-dimensional $I \in \mathbb{R}^{16}$, including 8 metrics for two images for comparison, half of which are ROI metrics. Our multilayer perceptron network structure is {192,64,64,32}, and the output is a bool used to point out the image with higher perceived quality. We use the cross entropy function as the loss function and the ADAM optimizer for training. After 50 iterations, the final results is shown in Table 1. Please note that for the base model, we only use the first row of metrics in Figure 1 for training and testing.

Table 1. Perceptual Metric

Metric	Accuracy
BASIC	76.72%
ROI	78.4%

References

[1] Zuyao Chen, Qianqian Xu, Runmin Cong, and Qingming Huang. Global context-aware progressive aggregation network for salient object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10599–10606, 2020. 1

	Image A				Image B			
BASIC	MSE	MS-SSIM	LPIPS	D-LOSS	MSE	MS-SSIM	LPIPS	D-LOSS
ROI	ROI MSE	ROI MS-SSIM	ROI LPIPS	ROI D-LOSS	ROI MSE	ROI MS-SSIM	ROI LPIPS	ROI D-LOSS

Figure 1. Input tensor $I \in \mathbb{R}^{16}$