# Perceptual Image Compression with Controllable Region Quality

Xiaohan Pan, Yixin Gao, Zongyu Guo, Runsen Feng, Zhibo Chen
University of Science and Technology of China
pxh123@mail.ustc.edu.cn, chenzhibo@ustc.edu.cn

## Abstract

*This one page factsheet describes our scheme in the track of image compression. We target at building an image compression model that supports controlling the perceptual quality of different regions. By generating image-adaptive quality maps from a saliency detection network, we can achieve better perceptual compression results at various image regions. Based on the network we used in CLIC2020 and CLIC2021, we add spatial feature transform (SFT) layers into the encoding and decoding transform networks, and use spatial-weighted loss for optimization to construct a bitrate controllable model. Our full compression model supports both variable-rate and controllable compression at different regions, while still delivering relatively high objective performance especially in terms of MS-SSIM.*

## 1. Introduction

In this short paper, we briefly introduce our scheme optimized for perceptual quality in the track of image compression. Our backbone is almost the same with our previous submission in CLIC2020 [1], which is designed to enhance the performance of network under the measure of objective indicators. It contains a causal 3D context model which leverages the dependencies along both spatial and channel dimensions in latent features, wherein a more accurate probability estimation model is achieved for entropy coding. We further replace the spatial-channel attention by group-separate attention module in our backbone, to strengthen the transform network. Based on this main network, we build a variable-rate image compression model with the aid of SFT, which is motivated from Song *et al*. [3].

Similar to the way described in [3], we insert the SFT layers into the autoencoder. The SFT layers take a quality map as the input, enabling convolutional networks to be aware of the parameters of element-wise affine transformation. With the help of such conditional transformation layers, the network achieves the property of controlling intermediate features within the encoding and decoding transform networks, which helps us build an image compression model with controllable bit rates. Different from the architecture in [3], we abandon the SFT layers inserted into hyper transform network, while still reach comparable performance with models trained for specific bit rate. Besides, by using a controllable map, we can obtain the ability of bit allocation in spatial dimension to some extent, which can be helpful when taking consideration of region of interest.

Perceptual-friendly image compression model can be built by applying perceptual loss and generative adversarial network (GAN) during training, such as the work of [2]. In our scheme, we adopt the similar strategy. We take the weighted MSE loss for spatially image quality control, and add the MS-SSIM loss to get sharper texture. The final loss item for training can be formulated as below:

$$\begin{aligned} L =& R + \sum_{i}^{N} \lambda_i \cdot \frac{(x_i - \hat{x}_i)}{N} + \lambda_M \cdot (1 - d_{MS-SSIM}) \\ &+ \lambda_L \cdot d_{LPIPS} + \lambda_G \cdot d_{GAN} \end{aligned} \tag{1}$$

where $R$, $d_{MS-SSIM}$, $d_{LPIPS}$, $d_{GAN}$ represent estimated bit rate and distortion metrics calculated by MS-SSIM, LPIPS and discriminator respectively. Here, $\lambda_i$ are pixel-wise weight generated from quality map through a predefined mapping function, as described in [3]. The other weights $\lambda_M$, $\lambda_L$, $\lambda_G$ are proportional to $\sum \lambda_i$.

Our model reaches control in bit-rate and region quality, while still maintain relatively high objective performance.

## References

[1] Zongyu Guo, Yaojun Wu, Runsen Feng, Zhizheng Zhang, and Zhibo Chen. 3-d context entropy model for improved practical image compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 116–117, 2020. 1

[2] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression. *Advances in Neural Information Processing Systems*, 33:11913–11924, 2020. 1

[3] Myungseo Song, Jinyoung Choi, and Bohyung Han. Variable-rate deep image compression through spatially-adaptive feature transform. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2380–2389, 2021. 1