Learning Scalable ℓ_∞-constrained Near-lossless Image Compression via Joint Lossy Image and Residual Compression
Supplementary Material

1. Lossy Image Compressor Architecture

As shown in Fig. 1, we employ the hyper-prior model proposed in [1], where side information $\hat{\mathbf{z}}$ is extracted to model the distribution of $\hat{\mathbf{y}}$. We employ the residual and attention blocks to model analysis/synthesis transforms, similar to [2, 3].

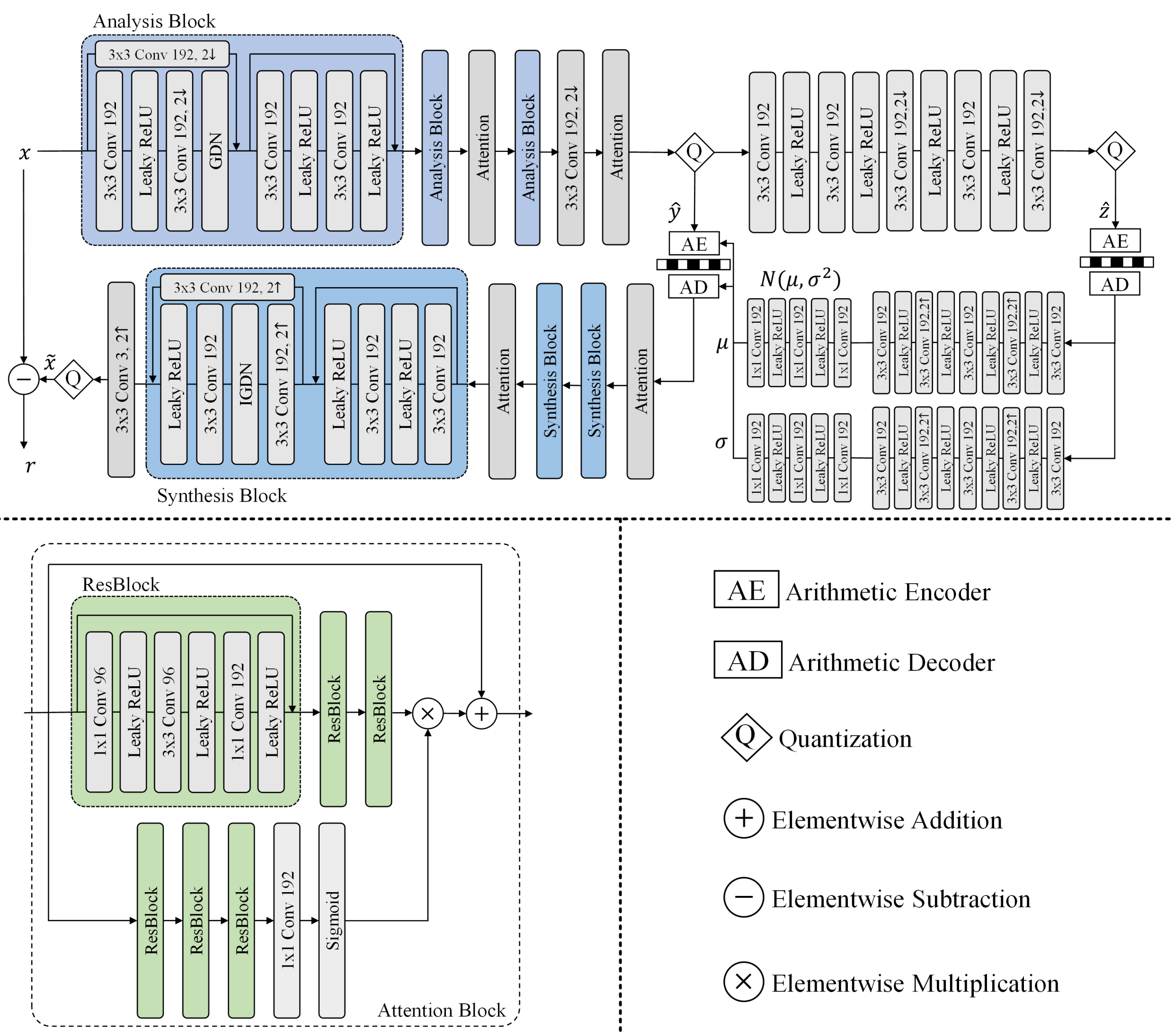


Figure 1. Lossy image compressor architecture. Top: overview. Bottom left: attention block. Bottom right: legend.

2. Proof of Entropy Reduction Resulting from Uniform Quantization

Proposition 1. *Let $x \in \{a, a+1, \dots, a+N\}$ be a discrete random variable with probability mass function (PMF) $p(x)$, where a and $N \geq 0$ are two integers. Suppose that x is quantized to $\hat{x}$ by*

$$\hat{x} = \mathrm{sgn}(x)(2\tau+1)\lfloor(|x|+\tau)/(2\tau+1)\rfloor, \quad \tau = 0, 1, \dots, M \tag{1}$$

where $\mathrm{sgn}(\cdot)$ denotes the sign function and $M \geq 0$ is an integer. The PMF $\hat{p}(\hat{x})$ of the quantized $\hat{x}$ is

$$\hat{p}(\hat{x}) = \sum_{x=\hat{x}-\tau}^{\hat{x}+\tau} p(x) \tag{2}$$

Let $H(x)$ and $H(\hat{x})$ be the entropy of x and $\hat{x}$ respectively, we have

$$H(x) \geq H(\hat{x}) \tag{3}$$

In words, uniformly quantizing x with (1) leads to entropy reduction.

Proof. Based on the definition of entropy and (2), we have

$$\begin{aligned} H(x) &= -\sum_{x} p(x) \log p(x) && (4) \\ &= -\sum_{\hat{x}} \sum_{x=\hat{x}-\tau}^{\hat{x}+\tau} p(x) \log p(x) && (5) \\ &\geq -\sum_{\hat{x}} \sum_{x=\hat{x}-\tau}^{\hat{x}+\tau} p(x) \log \hat{p}(\hat{x}) && (6) \\ &= -\sum_{\hat{x}} \hat{p}(\hat{x}) \log \hat{p}(\hat{x}) = H(\hat{x}) && (7) \end{aligned}$$

Inequation (6) holds, because $\hat{p}(\hat{x}) \geq p(x)$ for $x \in [\hat{x}-\tau, \hat{x}+\tau]$. □

Proposition 1 is the theoretical foundation for our scalable compression scheme corresponding to residual quantization.

3. More Visual Results

We show more visual results of the near-lossless reconstructions of our codec on Kodak dataset [4] in Fig. 2, 3, 4 and 5. Compared with the raw images, no visual artifacts are introduced by residual quantization. Human eyes can hardly differentiate between the the raw images and the near-lossless reconstructions of our codec at $\tau \leq 5$.

References

[1] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. In *Int. Conf. Learn. Represent.*, 2018. 1

[2] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7939–7948, 2020. 1

[3] Zongyu Guo, Yaojun Wu, Runsen Feng, Zhizheng Zhang, and Zhibo Chen. 3-d context entropy model for improved practical image compression. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 116–117, 2020. 1

[4] Eastman Kodak. Kodak lossless true color image suite (photocd pcd0992), 1993. `http://r0k.us/graphics/kodak/`. 2

(a) Raw image / $\lambda, \tau = 0.03, 0$

(b) $\lambda, \tau = 0.03, 1$

(c) $\lambda, \tau = 0.03, 2$

(d) $\lambda, \tau = 0.03, 3$

(e) $\lambda, \tau = 0.03, 4$

(f) $\lambda, \tau = 0.03, 5$

Figure 2. *Kodim09*. Near-lossless image codec is trained with $\lambda = 0.03$. Visualization of near-lossless reconstructions with different τ.

(a) Raw image / $\lambda, \tau = 0.03, 0$

(b) $\lambda, \tau = 0.03, 1$

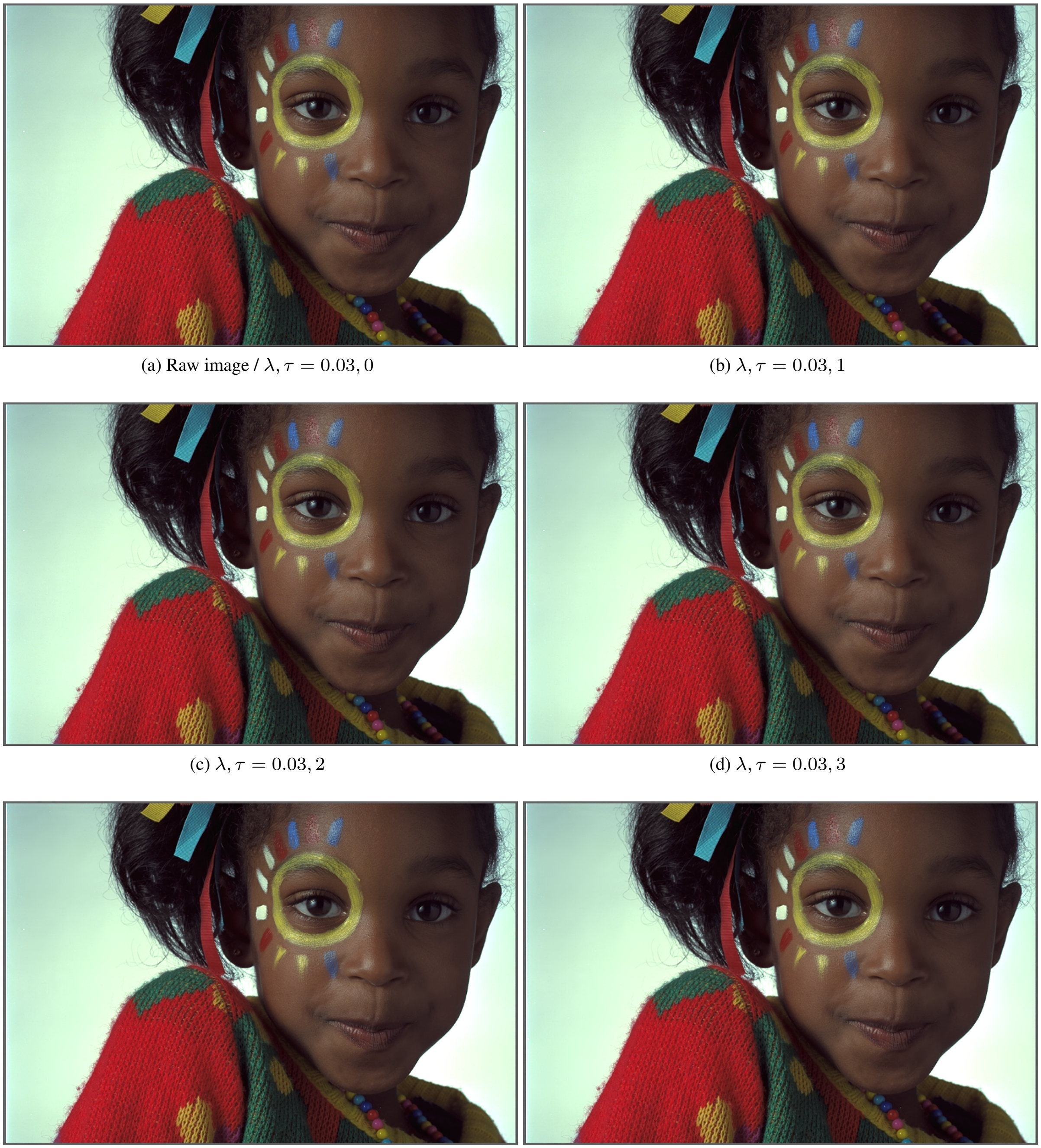
(c) $\lambda, \tau = 0.03, 2$

(d) $\lambda, \tau = 0.03, 3$

(e) $\lambda, \tau = 0.03, 4$

(f) $\lambda, \tau = 0.03, 5$

Figure 3. *Kodim13*. Near-lossless image codec is trained with $\lambda = 0.03$. Visualization of near-lossless reconstructions with different τ.



(a) Raw image / $\lambda, \tau = 0.03, 0$

(b) $\lambda, \tau = 0.03, 1$

(c) $\lambda, \tau = 0.03, 2$

(d) $\lambda, \tau = 0.03, 3$

(e) $\lambda, \tau = 0.03, 4$

(f) $\lambda, \tau = 0.03, 5$

Figure 4. *Kodim15*. Near-lossless image codec is trained with $\lambda = 0.03$. Visualization of near-lossless reconstructions with different τ.

(a) Raw image / $\lambda, \tau = 0.03, 0$

(b) $\lambda, \tau = 0.03, 1$

(c) $\lambda, \tau = 0.03, 2$

(d) $\lambda, \tau = 0.03, 3$

(e) $\lambda, \tau = 0.03, 4$

(f) $\lambda, \tau = 0.03, 5$

Figure 5. *Kodim21*. Near-lossless image codec is trained with $\lambda = 0.03$. Visualization of near-lossless reconstructions with different τ.