SUPPLEMENTARY MATERIAL FOR "AUTOENCODER BASED IMAGE COMPRESSION: CAN THE LEARNING BE QUANTIZATION INDEPENDENT?": ERROR OF ENTROPY APPROXIMATION

Thierry Dumas, Aline Roumy, Christine Guillemot

INRIA Rennes Bretagne-Atlantique
thierry.dumas@inria.fr, aline.roumy@inria.fr, christine.guillemot@inria.fr

1. INTRODUCTION

In [2], an approximation of the entropy of quantized transform coefficients is presented. The quantization is uniform scalar with a given quantization step size. As this approximation involves transform coefficients with additive signal-independent noise, the derivative of the approximate entropy of quantized transform coefficients with respect to a parameter of the parametrized encoder generating the transform coefficients is never zero. That is why it is useful in training the parametrized encoder.

This supplementary material compares this approximation of the entropy of quantized transform coefficients, called A.[2], with another approximation, called A.TH:8.3.1. To do so, the origin of the inaccuracies in both approximations are pointed out. Then, the two errors of entropy approximation are measured and compared.

2. APPROXIMATE ENTROPY A.TH:8.3.1

This section explains the approximation of the entropy of quantized transform coefficients, called A.TH:8.3.1, which is based on Theorem 8.3.1 in [1].

Let Y be a continuous random variable with probability density function p. In our case, transform coefficients are viewed as realizations of Y. The range of Y is divided into bins of length $\delta \in \mathbb{R}_+^*$. Using the mean value theorem, there exists a value $\hat{y}_i$ within each bin such that

$$\int_{i\delta}^{(i+1)\delta} p\left(y\right) dy = p\left(\hat{y}_i\right) \delta. \tag{1}$$

Let $\hat{Y}^{(0)}$ be the discrete random variable defined by quantizing Y via (2).

$$\hat{Y}^{(0)} = \hat{y}_i \text{ if } i\delta \leq Y < \left(i+1\right)\delta \tag{2}$$

The probability mass function $\hat{p}^{(0)}$ of $\hat{Y}^{(0)}$ is defined by

$$\hat{p}^{(0)}\left(\hat{y}_i\right) = \int_{i\delta}^{(i+1)\delta} p\left(y\right) dy = p\left(\hat{y}_i\right)\delta. \tag{3}$$

The entropy $H\left(\hat{Y}^{(0)}\right)$ of $\hat{Y}^{(0)}$ is expressed as

$$\begin{aligned} H\left(\hat{Y}^{(0)}\right) &= -\sum_i \hat{p}^{(0)}\left(\hat{y}_i\right) \log_2\left(\hat{p}^{(0)}\left(\hat{y}_i\right)\right) \\ &= -\sum_i \delta p\left(\hat{y}_i\right) \log_2\left(p\left(\hat{y}_i\right)\right) - \log_2\left(\delta\right). \end{aligned} \tag{4}$$

The quantization intervals together with the set $\{\hat{y}_i\}$ form a tagged partition of the range of Y. Using the approximation of the Riemann integral by the Riemann sums with respect to this tagged partition,

$$H\left(\hat{Y}^{(0)}\right) \rightarrow h\left(Y\right) - \log_2\left(\delta\right) \text{ as } \delta \rightarrow 0. \tag{5}$$

$h\left(Y\right) = -\int_y p\left(y\right) \log_2\left(p\left(y\right)\right) dy$ is the differential entropy of Y. But, in the context of compression, the uniform scalar quantization with quantization step size δ is used instead of the quantization defined in (2). Let $\hat{Y}^{(1)}$ be the discrete random variable defined by quantizing Y via this uniform scalar quantization.

$$\hat{Y}^{(1)} = i\delta + 0.5\delta \text{ if } i\delta \leq Y < \left(i+1\right)\delta \tag{6}$$

As this uniform scalar quantization and the quantization in (2) get close as δ tends to 0,

$$H\left(\hat{Y}^{(1)}\right) \rightarrow h\left(Y\right) - \log_2\left(\delta\right) \text{ as } \delta \rightarrow 0. \tag{7}$$

Therefore, in the derivation of (7), we identify two sources of inaccuracy for the approximation of $H\left(\hat{Y}^{(1)}\right)$.

- The approximation of the Riemann integral by the Riemann sums.
- The use of the uniform scalar quantization instead of the quantization in (2).

Now, using n realizations $\{s_1, ..., s_n\}$ of Y, the differential entropy $h\left(Y\right)$ of Y is estimated, and (7) becomes

$$H\left(\hat{Y}^{(1)}\right) \approx -\frac{1}{n}\sum_{j=1}^{n} \log_2\left(p\left(s_j\right)\right) - \log_2\left(\delta\right). \tag{8}$$

The approximation A.TH:8.3.1 to be compared in Section 4 corresponds to (8).

3. APPROXIMATE ENTROPY A.[2]

This section recalls the approximation of the entropy of quantized transform coefficients, called A.[2], in [2].

Let $\{\tau_1, ..., \tau_n\}$ be n samples from the continuous uniform distribution of support $[-0.5, 0.5]$. Then, A.[2] is expressed as

$$H\left(\hat{Y}^{(1)}\right) \approx -\frac{1}{n}\sum_{j=1}^{n} \log_2\left(\tilde{p}\left(s_j + \delta\tau_j\right)\right) - \log_2\left(\delta\right). \quad (9)$$

$\tilde{p} = p * l$ where l is the probability density function of the continuous uniform distribution of support $[-0.5\delta, 0.5\delta]$.

While the approximation of the entropy of quantized transform coefficients given by (8) contains two sources of inacurracy, the approximation in (9) has only one source of inaccuracy: the noisy transform coefficients $\{s_j + \delta\tau_j\}$ replace the quantized transform coefficients.

4. ERRORS OF ENTROPY APPROXIMATION

Section 4 compares the error of entropy approximation given by A.TH:8.3.1 and that of A.[2]. The numerical results below can be reproduced via the script "svhn/comparing_approximations_entropy.py".

For A.TH:8.3.1 and A.[2], the two errors of entropy approximation are the absolute difference between the left term and the right term in (8) and (9) respectively. From now on, $n = 200000$. $\tilde{p}$ is approximated by fitting a parametrized function to it. This fitting is re-run every time either p or δ changes. Note that, the number of samples n is so large that, in (8), the estimation of the differential entropy $h\left(Y\right)$ of Y can be replaced by the true differential entropy of the known distribution below.

4.1. Normal distribution

Y follows the normal distribution of mean 0.0 and standard deviation 3.0, denoted $\mathcal{N}\left(0.0, 3.0\right)$. In this case, Figures 1 compares the evolution of the two errors of entropy approximation with δ.

4.2. Logistic distribution

Y follows the logistic distribution of mean 0.0 and scale 1.0, denoted $\mathcal{L}\left(0.0, 1.0\right)$. In this case, Figure 2 compares the evolution of the two errors of entropy approximation with δ.

4.3. Laplace distributions

Y follows the Laplace distribution of mean 0.0 and scale $\lambda \in \mathbb{R}_+^*$, denoted $L\left(0.0, \lambda\right)$. In this case, Figures 3, 4, and 5 compare the evolution of the two errors of entropy approximation with δ when $\lambda = 0.5$, $\lambda = 1.0$, and $\lambda = 2.0$ respectively.

Fig. 1: Evolution of the two errors of entropy approximation with δ when $Y \sim \mathcal{N}\left(0.0, 3.0\right)$

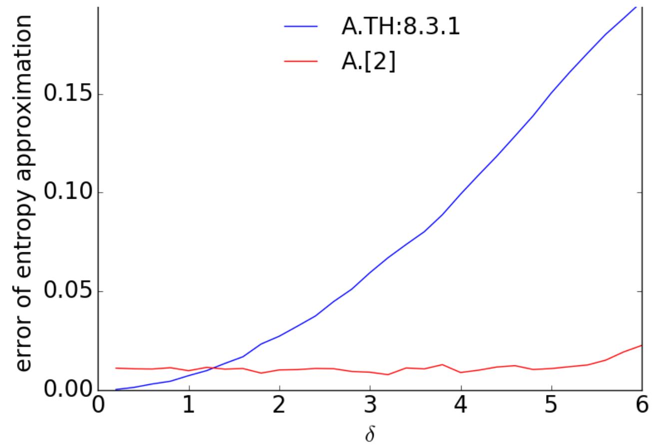
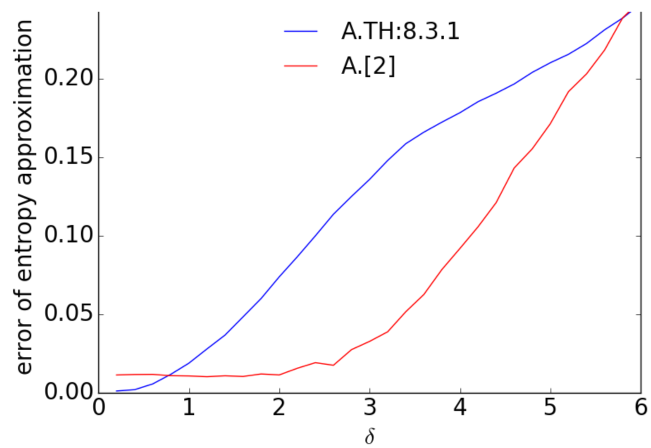


Fig. 2: Evolution of the two errors of entropy approximation with δ when $Y \sim \mathcal{L}\left(0.0, 1.0\right)$



5. CONCLUSION

A.[2] provides a more precise approximation of the entropy of quantized transform coefficients than A.TH:8.3.1 when $\delta \in [1.0, 6.0]$.

6. REFERENCES

[1] Thomas M. Cover and Joy A. Thomas. *Elements of information theory, second edition*. Wiley intersciences, 2006.

[2] Thierry Dumas, Aline Roumy, and Christine Guillemot. Autoencoder based image compression: can the learning be quantization independent? In *ICASSP*, 2018.

Fig. 3: Evolution of the two errors of entropy approximation with δ when $Y \sim L\,(0.0, 0.5)$

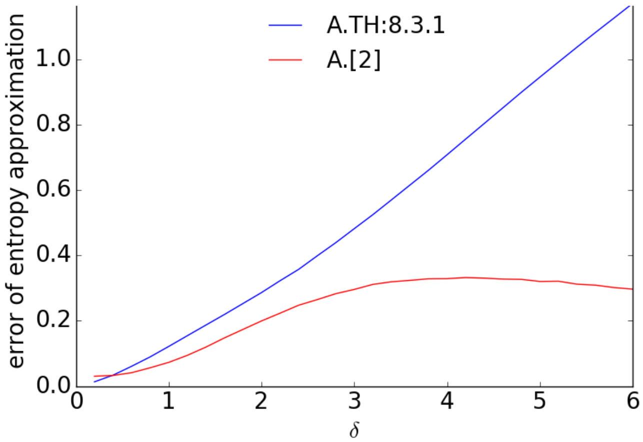


Fig. 4: Evolution of the two errors of entropy approximation with δ when $Y \sim L\,(0.0, 1.0)$

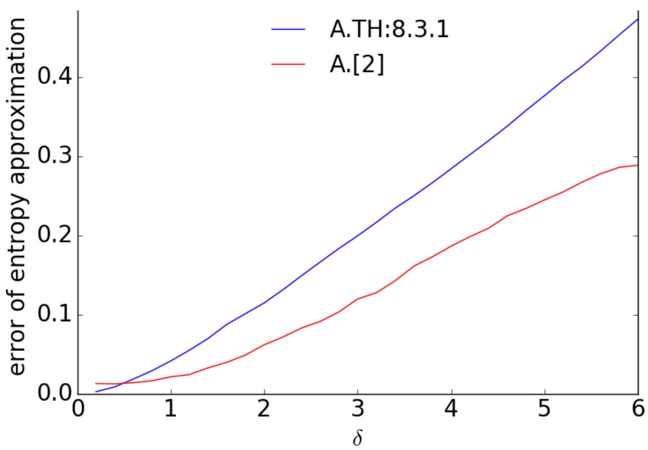


Fig. 5: Evolution of the two errors of entropy approximation with δ when $Y \sim L\,(0.0, 2.0)$