Binary Probability Model for Learning Based Image Compression

Paper ID: 1415

Théo LADUNE[1,2], Pierrick PHILIPPE[1], Wassim HAMIDOUCHE[2], Lu ZHANG[2], Olivier DÉFORGES[2]

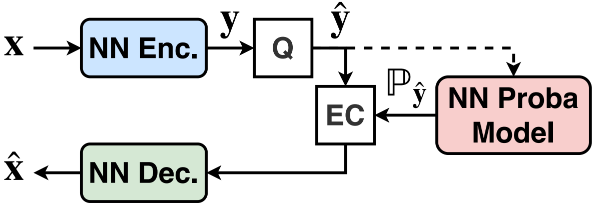
[1]Orange Labs, France — [2]INSA Rennes, France
`theo.ladune@orange.com`

International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2020

Introduction

- Context
  - Learned image coding competitive is now with **BPG** (HEVC-based image coding method)
  - Improvements mainly due to a better estimate of the **latents probability distribution** leading to better **entropy coding**
- Purpose
  - Propose a **richer probability distribution** to better model the latents
  - Improve **entropy coding** performances
- Results
  - **-9 % rate** in comparison to gaussian-based state-of-the-art systems
  - Perform on par with state-of-the-art systems with around **10x less weights**

- Learned Image Compression
  1. **Encoding** input image $\mathbf{x}$ into (quantized) latents $\hat{\mathbf{y}}$
  2. **Entropy Coding** (EC) with a probability model $\mathbb{P}_{\hat{\mathbf{y}}}$
  3. **Decoding** $\hat{\mathbf{y}}$ to reconstruct input image $\hat{\mathbf{x}}$



- **Rate-distortion** loss function

$$\mathcal{L}(\lambda) = \mathrm{D}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \mathrm{R}(\hat{\mathbf{y}}).$$

Entropy Coding

- Latents $\hat{\mathbf{y}}$ are sent using an **entropy coding** method, their rate is

$$\mathrm{R}(\hat{\mathbf{y}}) = \overbrace{\mathbb{E}_{\hat{\mathbf{y}} \sim m}}^{\text{Unknown distribution}}[-\log_2 \overbrace{\mathbb{P}_{\hat{\mathbf{y}}}(\hat{\mathbf{y}})}^{\text{Probability model}}]$$

- Rate is the cross entropy between an **unknown** distribution m and the probability model $\mathbb{P}_{\hat{\mathbf{y}}}$.

$$\mathrm{R}(\hat{\mathbf{y}}) = \mathrm{H}(m,\ \mathbb{P}_{\hat{\mathbf{y}}}) = \overbrace{\mathrm{H}(m)}^{\text{Encoder}} + \overbrace{D_{KL}(m \,||\, \mathbb{P}_{\hat{\mathbf{y}}})}^{\text{Probability model}} \geq \mathrm{H}(m)$$

- This work aims to **lower** $D_{KL}(m \,||\, \mathbb{P}_{\hat{\mathbf{y}}})$ through a more suited $\mathbb{P}_{\hat{\mathbf{y}}}$

Entropy coding

- Previous works assume **independence** for each latent

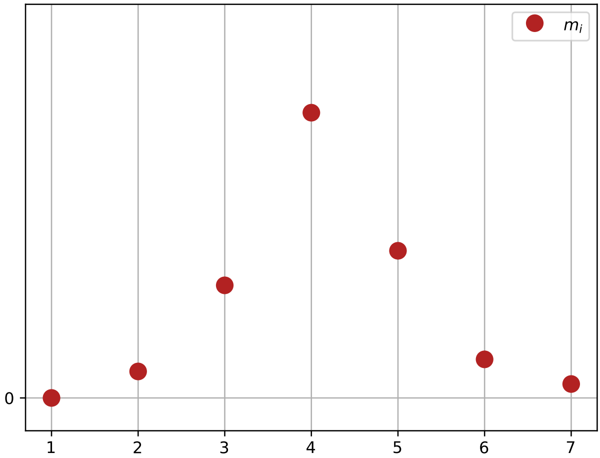
$$\mathbb{P}_{\hat{\mathbf{y}}}(\hat{\mathbf{y}}) = \prod_i \mathbb{P}_{\hat{y}_i}(\hat{y}_i)$$

- Each $\mathbb{P}_{\hat{y}_i}$ results from **a Gaussian[1,2] or Laplace[3] distribution**

- Each latent PDF parameters $\boldsymbol{\Psi}_i = \{\mu_i, \sigma_i\}$ are decoded from side-information and/or previously received latents.

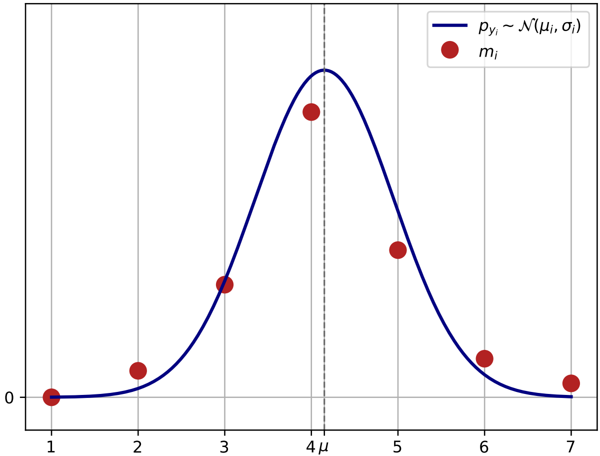
[1] Ballé, et al., Variational image compression with a scale hyperprior, ICLR 2018
[2] Minnen, et al., Joint Autoregressive and Hierarchical Priors for Learned Image Compression, NeurIPS 2018
[3] Zhou, et al., Variational autoencoder for low bit-rate image compression, CVPR 2018

Previous works



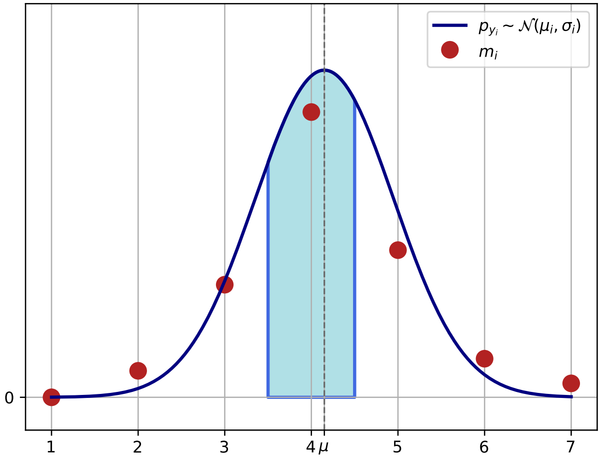
Underlying distribution m_i of the i-th latent

Previous works



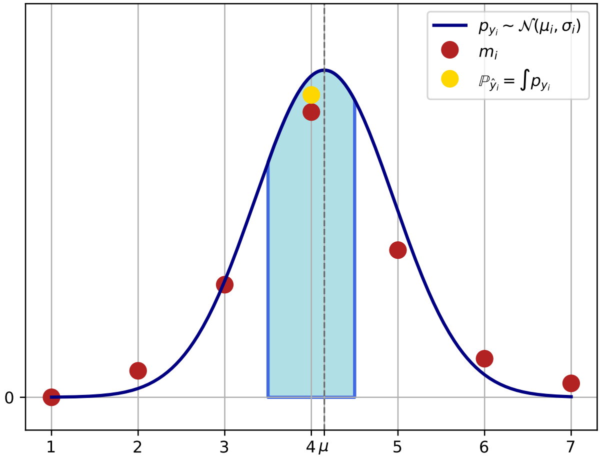
$\mathbb{P}_{\hat{y}_i}$ is modeled through p_{y_i}, a gaussian PDF for y_i

Previous works



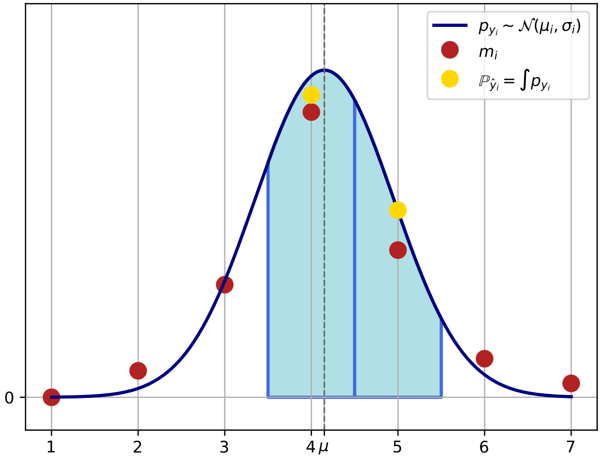
p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



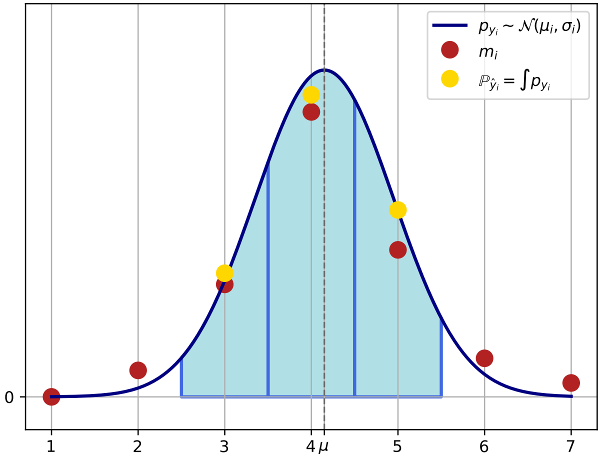
p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



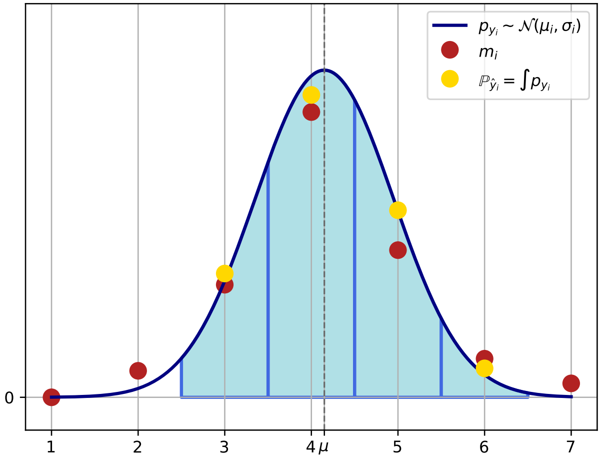
p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



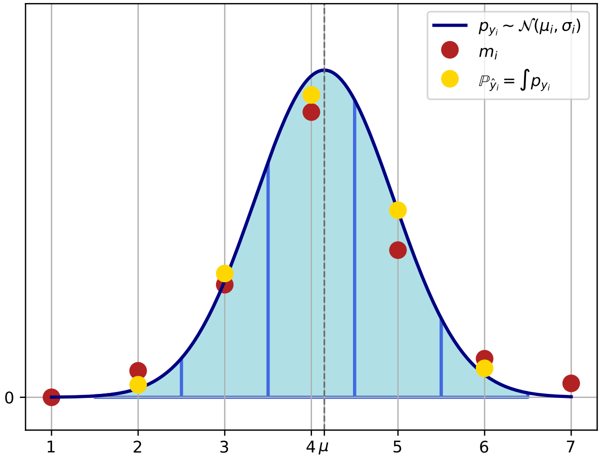
p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



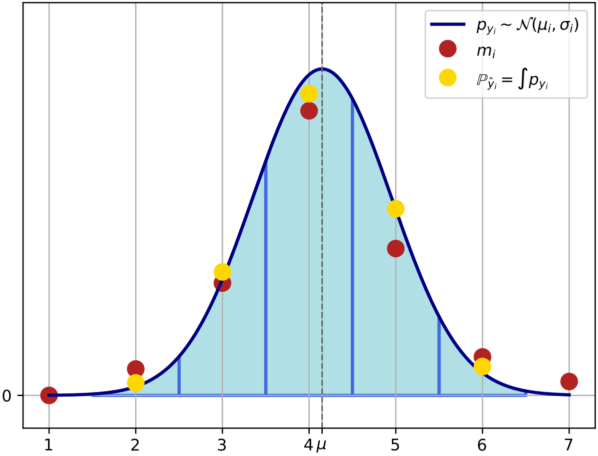
p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



p_{y_i} is integrated in each quantization bin to obtain $\mathbb{P}_{\hat{y}_i}$

Previous works



- Mismatch $D_{KL}(m \,||\, \mathbb{P}_{\hat{\mathbf{y}}})$ is important
- Rate $\mathrm{R}(\hat{\mathbf{y}}) = \mathrm{H}(m) + D_{KL}(m \,||\, \mathbb{P}_{\hat{\mathbf{y}}})$ is high

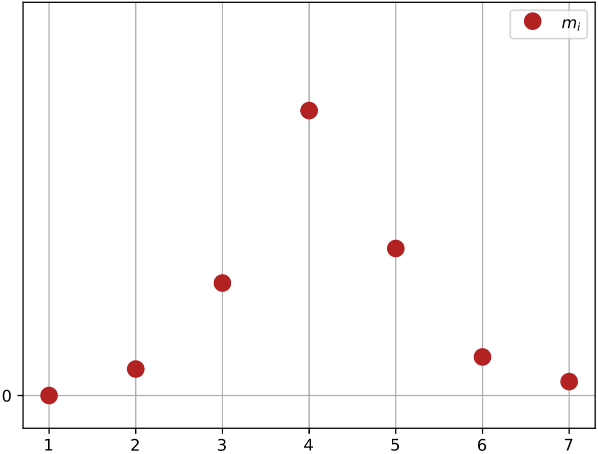
Binary Probability Model

Binary Probability Model:

- **Relax the entropy model** to reduce the mismatch $D_{KL}(m \mid\mid \mathbb{P}_{\hat{\mathbf{y}}})$
- As in classical video codecs, **binary flags** are used to freely set the probability of the 3 most frequent quantization bins (μ, $\mu - 1$, $\mu + 1$)
- This allows to represent **all symetrical distributions** in the interval $[\mu - 1, \mu + 1]$
- Rely on a **parametric PDF** (Laplace distribution) for all other quantization bins

Binary Probability Model

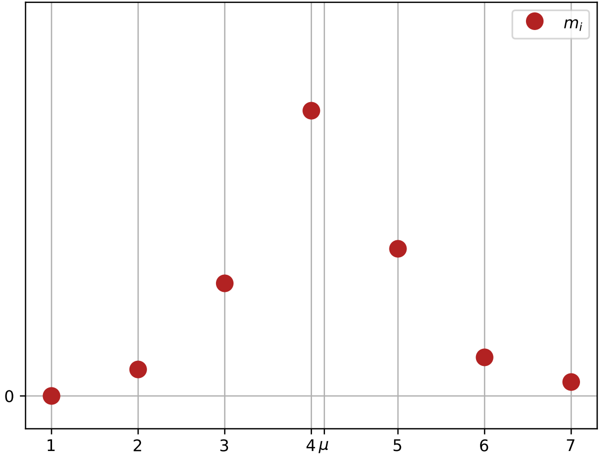
- Required parameters:



Underlying distribution m_i of the i-th latent

Binary Probability Model

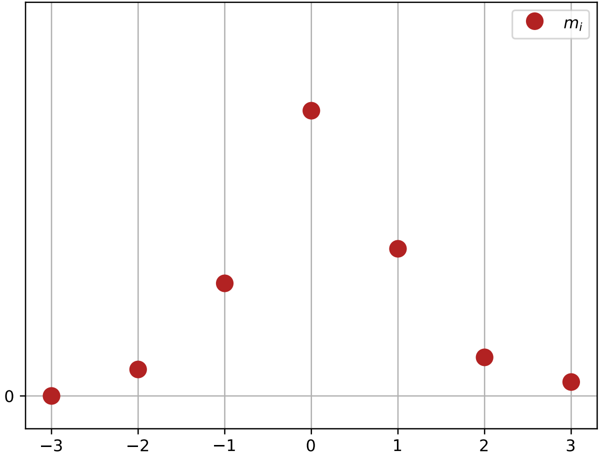
- Required parameters: μ_i



μ_i is used before quantization to center latent: $\hat{y}_i = Q(y_i - \mu_i)$

Binary Probability Model

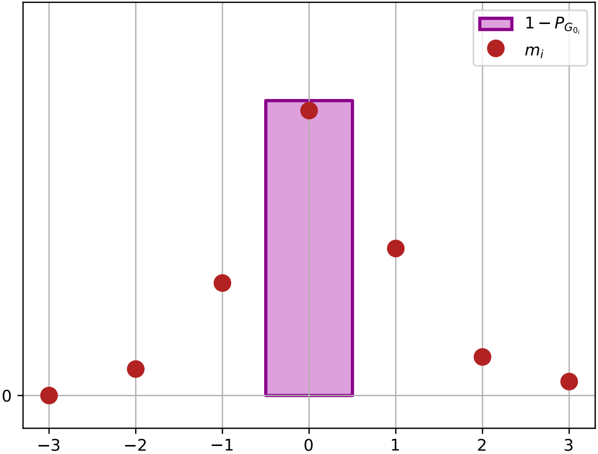
- Required parameters: μ_i



μ_i is used before quantization to center latent: $\hat{y}_i = Q(y_i - \mu_i)$

Binary Probability Model

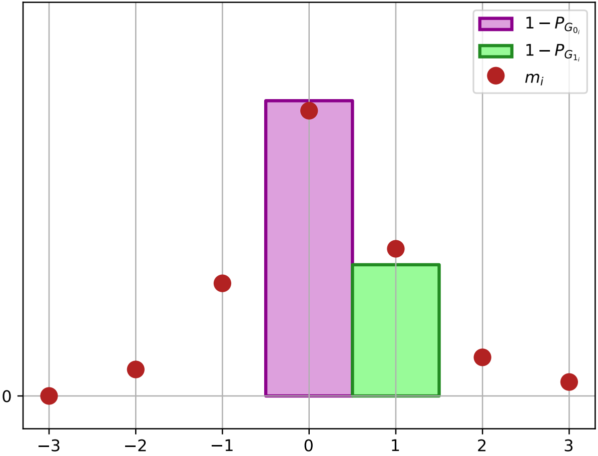
- Required parameters: μ_i, $P_{G_{0_i}}$



Probability $(1 - P_{G_{0_i}})$ of being in the 1st quantization bin is explicitely used

Binary Probability Model

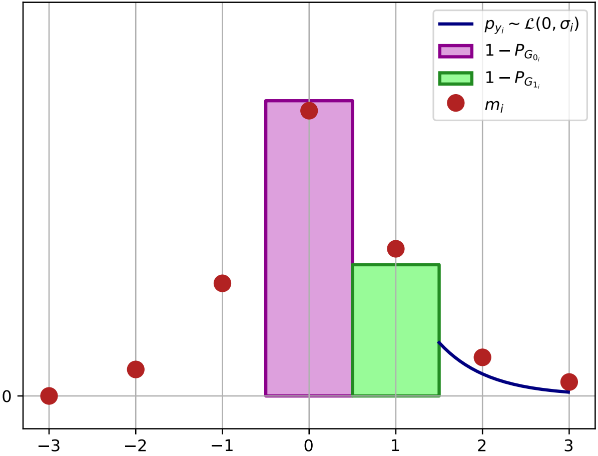
- Required parameters: μ_i, $P_{G_{0_i}}$, $P_{G_{1_i}}$



Probability $(1 - P_{G_{1_i}})$ of being in the 2^{nd} quantization bin is explicitely used

Binary Probability Model

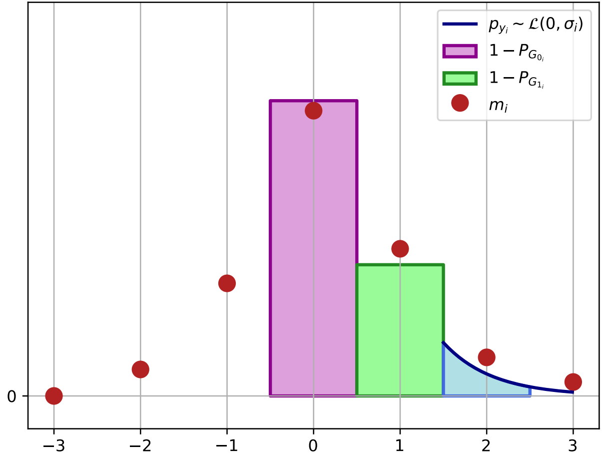
- Required parameters: μ_i, $P_{G_{0_i}}$, $P_{G_{1_i}}$, σ_i



Other quantization bins are modeled as in previous works using $\mathcal{L}(0, \sigma_i)$

Binary Probability Model

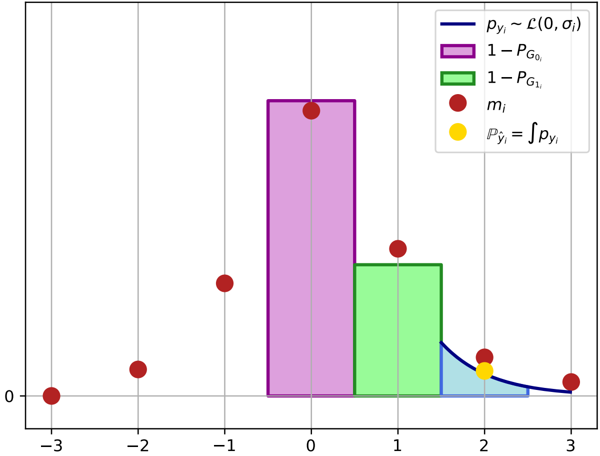
- Required parameters: μ_i, $P_{G_{0_i}}$, $P_{G_{1_i}}$, σ_i



Other quantization bins are modeled as in previous works using $\mathcal{L}(0, \sigma_i)$

Binary Probability Model

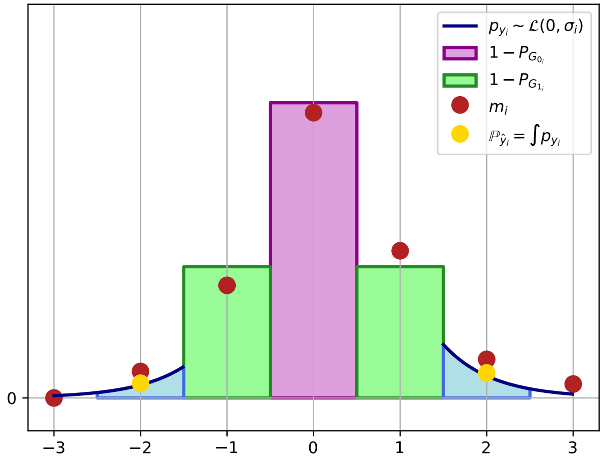
- Required parameters: μ_i, $P_{G_{0_i}}$, $P_{G_{1_i}}$, σ_i



Other quantization bins are modeled as in previous works using $\mathcal{L}(0, \sigma_i)$

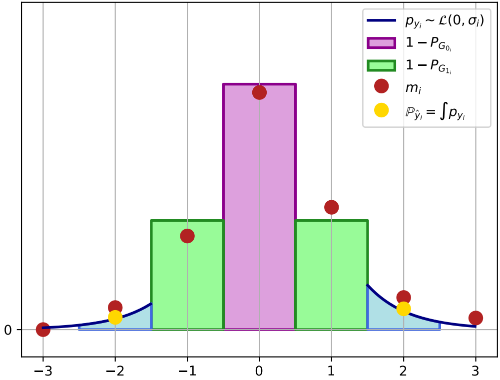
Binary Probability Model

- Required parameters: μ_i, $P_{G_{0_i}}$, $P_{G_{1_i}}$, σ_i



Symetrical distribution is assumed: $\mathbb{P}_{\hat{y}_i}(k) = \mathbb{P}_{\hat{y}_i}(-k)$

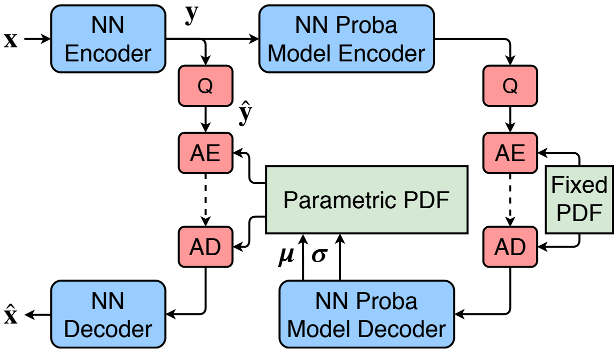
Binary Probability Model



- This example is for one latent, the same is done **for all latents**
- The most frequent quantization bin probabilities are **explicitely** set
- This allows to **better fit** $m \rightarrow$ Reduce $D_{KL}(m \mid\mid \mathbb{P}_{\hat{\mathbf{y}}})$
- Rate $\mathrm{R}(\hat{\mathbf{y}}) = \mathrm{H}(m) + D_{KL}(m \mid\mid \mathbb{P}_{\hat{\mathbf{y}}})$ **is lower**

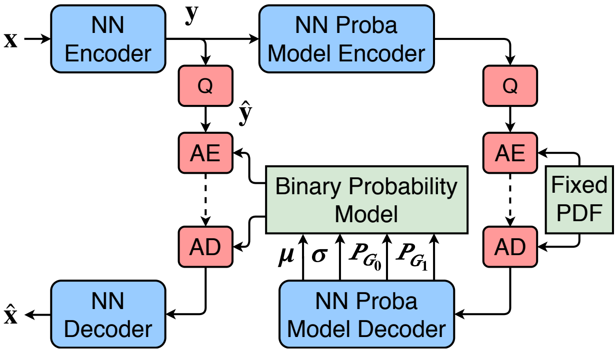
Network architecture

- The proposed binary probability model is implemented using the **exact same** architecture than previous works



Parametric PDF systems

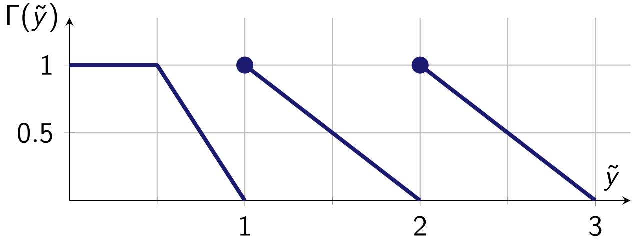
- The proposed binary probability model is implemented using the **exact same** architecture than previous works



Binary Probability Model systems – 2 supplementary parameters to convey

- During training, **continuous** $\tilde{y}$ replaces discrete $\hat{y}$
- Since R uses P_{G_0} and P_{G_1}, it is only defined for integers. An **interpolation** $\tilde{\mathrm{R}}$ is designed

$$\tilde{\mathrm{R}}(\tilde{y}) = \alpha \mathrm{R}(\lfloor \tilde{y} \rfloor) + (1-\alpha)\,\mathrm{R}(\lfloor \tilde{y} \rfloor + 1),\ \alpha = \Gamma(|\tilde{y}|)$$

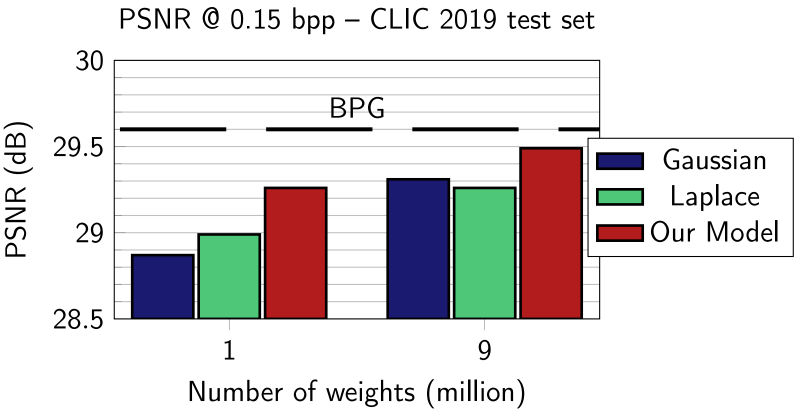


Experimental Results – Test conditions

- The proposed probability model is evaluated under the *Challenge on Learned Image Compression* **CLIC 2019 test conditions**
  - 330 images
  - Best PSNR at 0.15 bits per pixel (bpp)
- The proposed binary probability model is **compared to 2 parametric models**: a Gaussian[1,2] and Laplace[3] distributions
- All probability models are assessed in **2 configurations**
  - Lightweight (Around 1 million weights)
  - Standard (Around 9 million weights)

[1] Ballé, et al., Variational image compression with a scale hyperprior, ICLR 2018
[2] Minnen, et al., Joint Autoregressive and Hierarchical Priors for Learned Image Compression, NeurIPS 2018
[3] Zhou, et al., Variational autoencoder for low bit-rate image compression, CVPR 2018

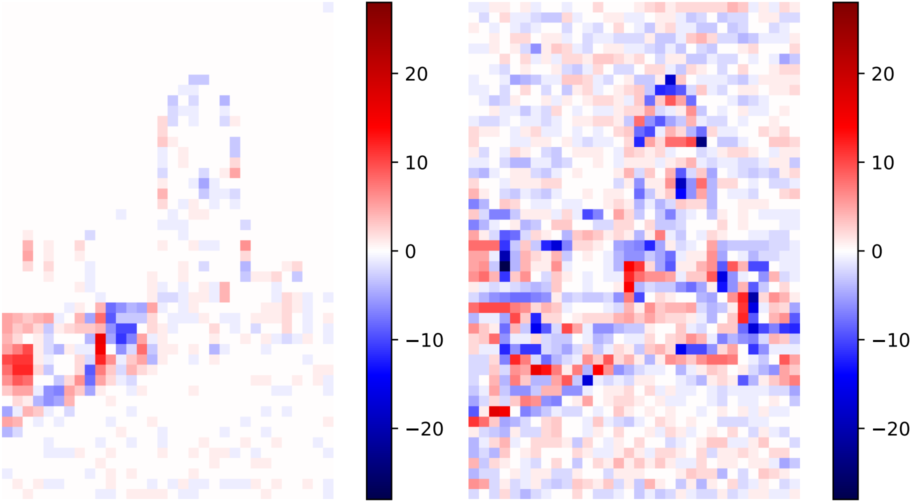
Experimental Results – Performances



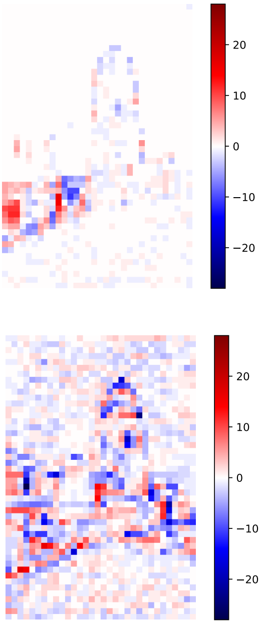
- PSNR 0.2 dB higher ↔ **-9 % rate** for the same quality compared to Gaussian systems
- Same performance than Gaussian systems with almost **10 times less weights**

Illustrations

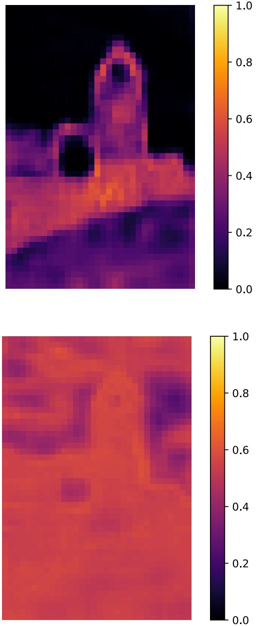
Input image to be coded



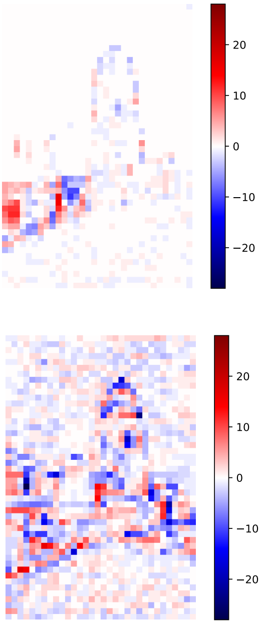
Two feature maps $\hat{\mathbf{y}}$ *to entropy code*



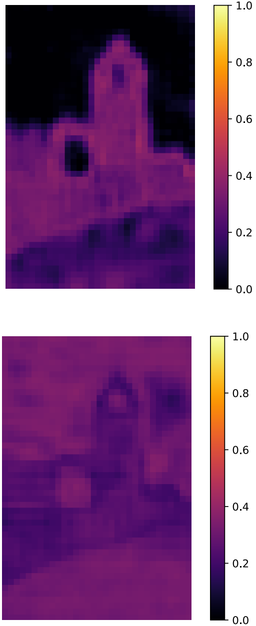
Latent feature maps $\hat{\mathbf{y}}$



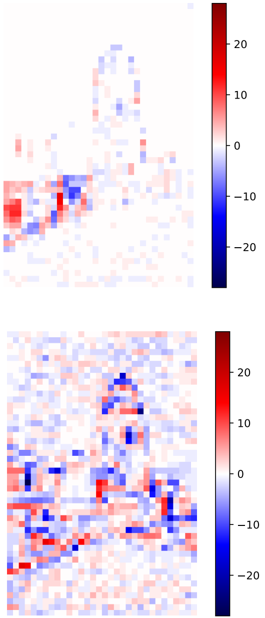
Probability of greater than 0 $\boldsymbol{P}_{\boldsymbol{G_0}}$



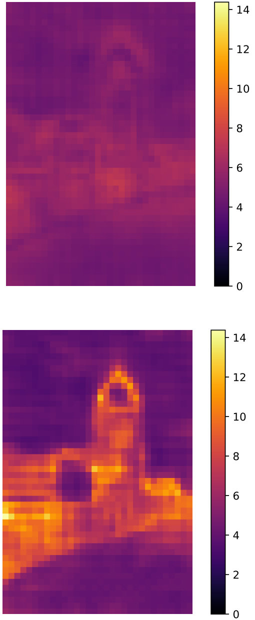
Latent feature maps $\hat{\mathbf{y}}$



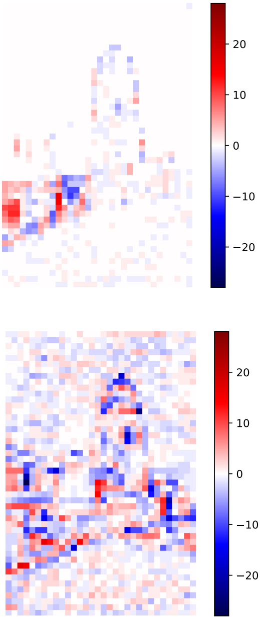
Probability of greater than 1 $\boldsymbol{P}_{\boldsymbol{G}_1}$



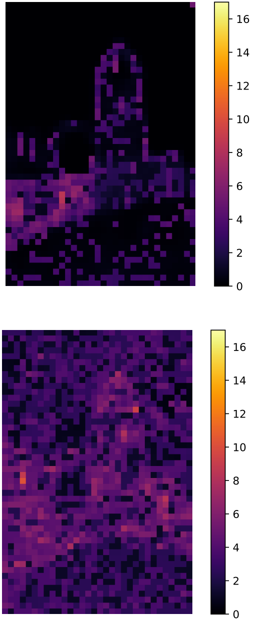
Latent feature maps $\hat{\mathbf{y}}$



$\boldsymbol{\sigma}$ *for Laplace distribution*



Latent feature maps $\hat{\mathbf{y}}$



Rate in bits

Conclusion

- We propose to **reduce the mismatch** between the unknown latents distribution and the probability model used for entropy coding
- This is achieved by **adding two parameters** to previous probability model: P_{G_0} and P_{G_1}
- This results in a more flexible probability model and a **better entropy coding**
- **-9 % rate** compared to a state-of-the-art gaussian model
- Binary Probability Model achieves performances competitive with gaussian model while using **10x less weights**

Binary Probability Model for Learning Based Image Compression

Paper ID: 1415

Théo LADUNE[1,2], Pierrick PHILIPPE[1], Wassim HAMIDOUCHE[2], Lu ZHANG[2], Olivier DÉFORGES[2]

[1]Orange Labs, France — [2]INSA Rennes, France
theo.ladune@orange.com

International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2020