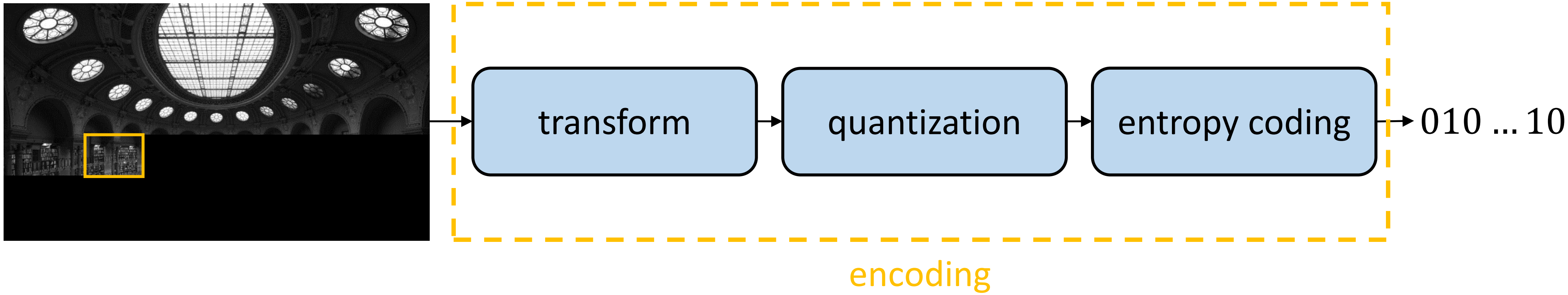
Autoencoder based image compression: can the learning be quantization independent?

Thierry Dumas, Aline Roumy, Christine Guillemot

INRIA Rennes, France

Transform coding



transform : KLT

- if image pixels ~ Gaussian, optimal
- not image independent
  ↳ need to transmit KLT basis

DCT

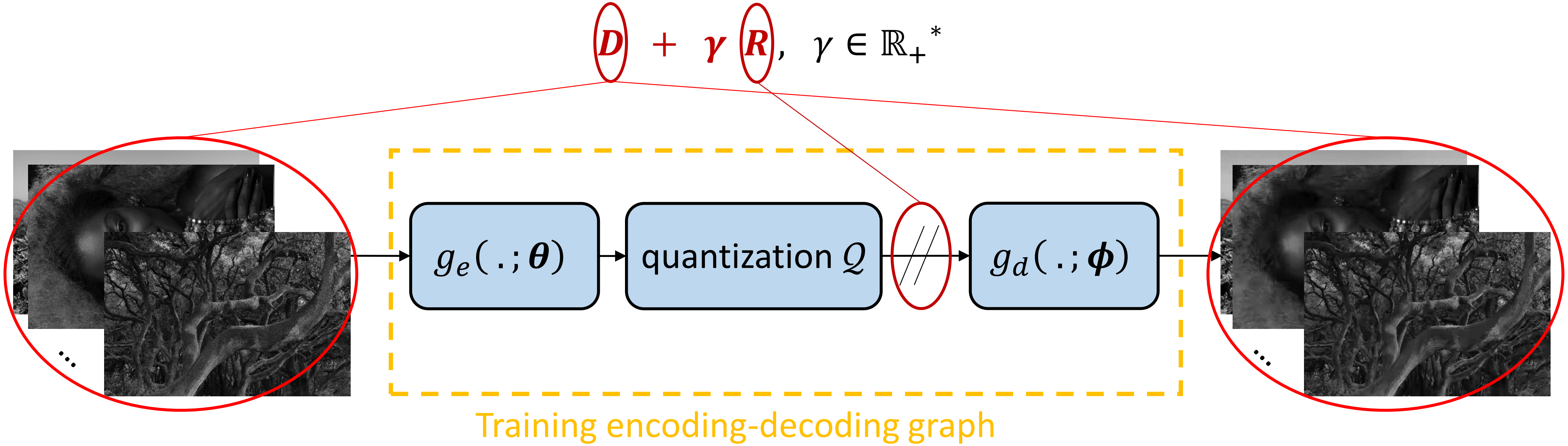
- if images ~ highly correlated GM process, almost optimal
- image independent
  ↳ no need to transmit DCT basis

Learning alternative transforms

- Discovering **alternative transforms**
- Suitable **for image compression**

➡ **learning** with **rate-distortion optimization**.

$$\boldsymbol{D} + \boldsymbol{\gamma}\boldsymbol{R}, \quad \gamma \in \mathbb{R}_+^*$$



From classical to universal

Usually,

Training

$\gamma_0 \longrightarrow \{\boldsymbol{\theta}_0{}^*, \boldsymbol{\phi}_0{}^*\}$

$\gamma_1 \longrightarrow \{\boldsymbol{\theta}_1{}^*, \boldsymbol{\phi}_1{}^*\}$

$\vdots \qquad \vdots$

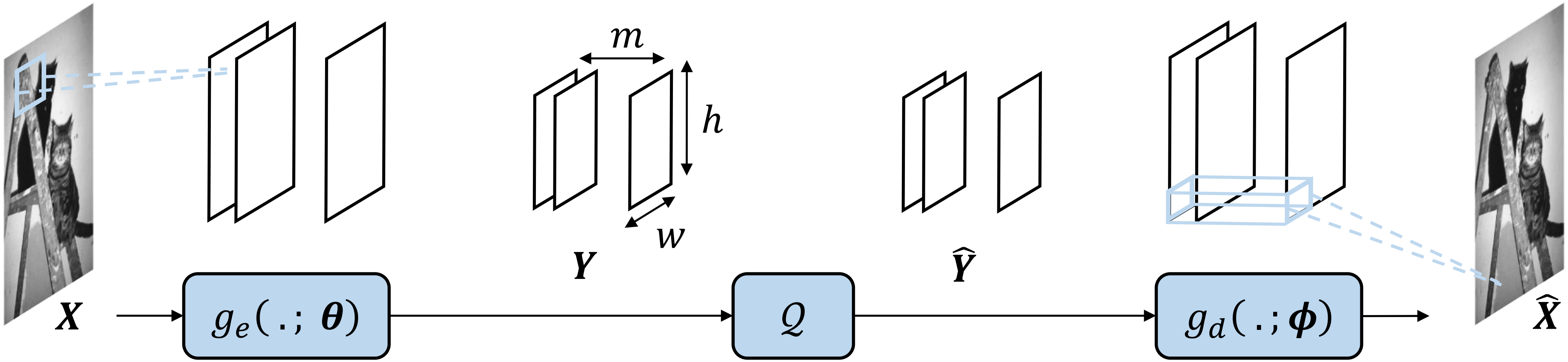
$\gamma_{n-1} \longrightarrow \{\boldsymbol{\theta}_{n-1}{}^*, \boldsymbol{\phi}_{n-1}{}^*\}$

Test

$$\left.\begin{matrix} \{\boldsymbol{\theta}_0{}^*, \boldsymbol{\phi}_0{}^*\} \\ \{\boldsymbol{\theta}_1{}^*, \boldsymbol{\phi}_1{}^*\} \\ \vdots \\ \{\boldsymbol{\theta}_{n-1}{}^*, \boldsymbol{\phi}_{n-1}{}^*\} \end{matrix}\right\} \longrightarrow$$ n rates for compression

- **Varying the quantization at test time**? ⟶ **Learning** and storing **one transform** instead of n.
- **Learning jointly the quantization and the transform**? ⟶ Towards optimality?

I – Autoencoder for image compression

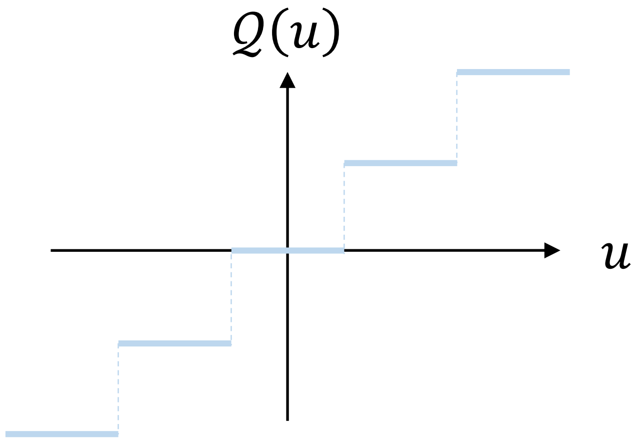


empirical entropy of the i^{th} feature map in $\boldsymbol{Y}$

$$\min_{\boldsymbol{\theta},\boldsymbol{\phi}} \underbrace{\mathbb{E}\left[\left\|\boldsymbol{X} - g_d\big(\mathcal{Q}\big(g_e(\boldsymbol{X};\boldsymbol{\theta})\big);\boldsymbol{\phi}\big)\right\|_2^{\,2}\right]}_{D} + \gamma\,\underbrace{\mathbb{E}\left[\sum_{i=1}^{m} -\frac{1}{h\times w}\sum_{j=1}^{h\times w}\log_2\Big(\hat{p}_i\big(\hat{y}_{ij}\big)\Big)\right]}_{R}$$

$$D + \gamma\, R$$

I – Autoencoder for image compression



$\mathcal{Q}'(u) = 0$ ➡ ***θ* cannot be learned** via gradient-based methods

⬇

approximating $\mathcal{Q}$ at training time.

II – Two learnings

$\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_m\}, \delta_i$ = quantization step size for $\mathcal{Q}_i, i \in [\![1, m]\!]$.

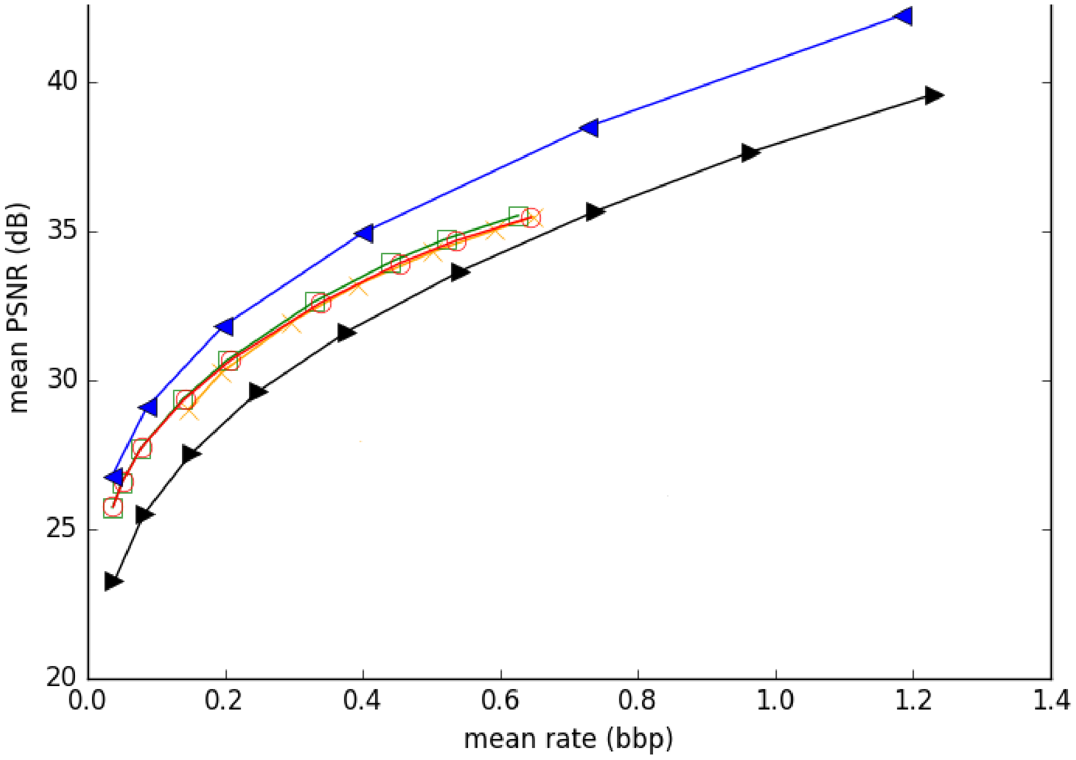
- Learning jointly $\{\boldsymbol{\theta}, \boldsymbol{\phi}, \delta_1, \ldots, \delta_m\}$:

noise $\mathcal{U}[-0.5, 0.5]$

$$\min_{\boldsymbol{\theta}, \boldsymbol{\phi}, \delta_1, \ldots, \delta_m} \mathbb{E}\left[\left\| \boldsymbol{X} - g_d\left(g_e(\boldsymbol{X}; \boldsymbol{\theta}) + \boldsymbol{\Delta} \odot \boldsymbol{T}; \boldsymbol{\phi}\right) \right\|_F^2 \right] + \gamma\, \mathbb{E}\left[\sum_{i=1}^{m} \left(-\log_2(\delta_i) - \frac{1}{h \times w} \sum_{j=1}^{h \times w} \log_2\left(\tilde{p}_i\left(y_{ij} + \delta_i \tau_{ij} \right) \right) \right) \right]$$

- Learning $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ while fixing $\{\delta_1, \ldots, \delta_m\}$.

III - Experiments



H.265

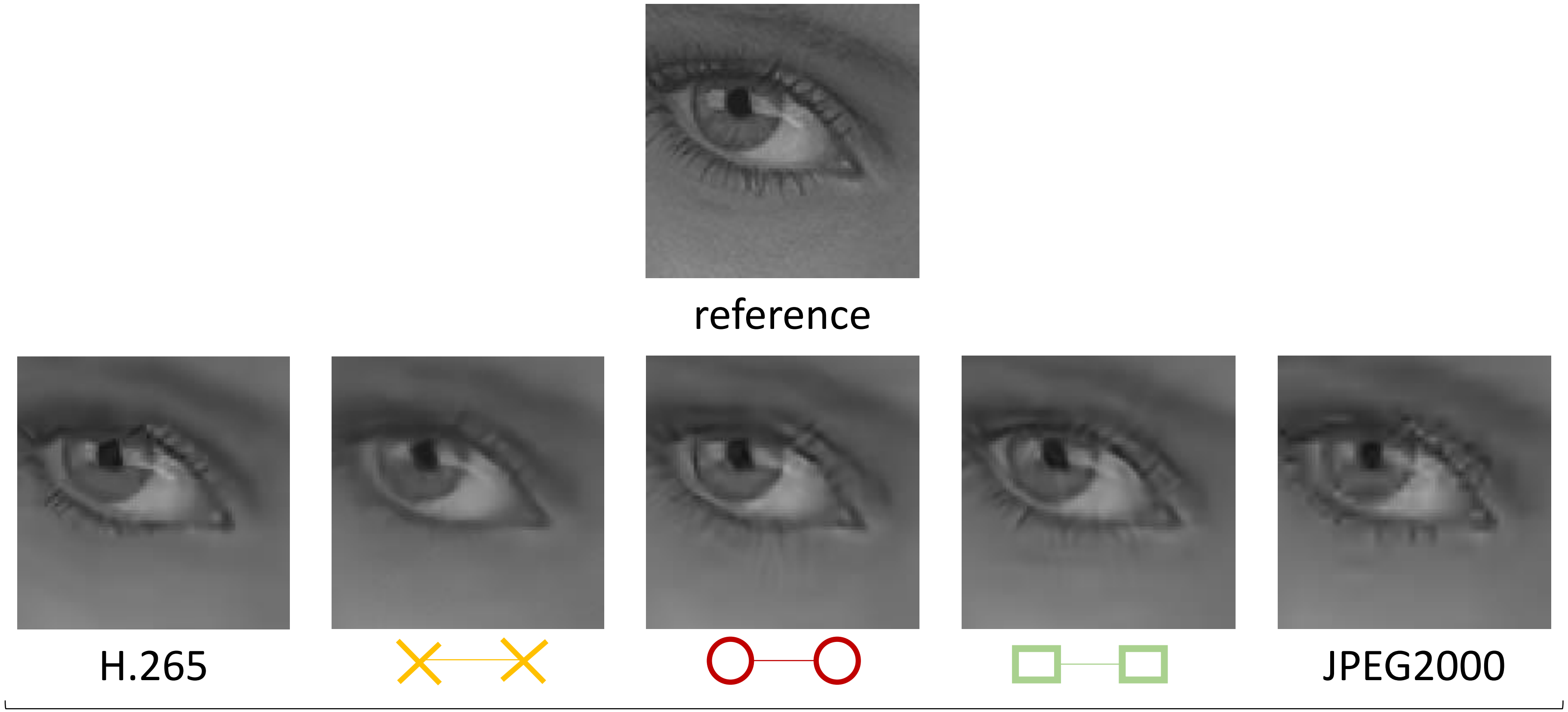
- learning $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ while fixing $\{\delta_1, \ldots, \delta_m\}$; one $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ per γ
- still fixing $\{\delta_1, \ldots, \delta_m\}$ at test time

- learning $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ while fixing $\{\delta_1, \ldots, \delta_m\}$
- varying $\{\delta_1, \ldots, \delta_m\}$ at test time

- learning jointly $\{\boldsymbol{\theta}, \boldsymbol{\phi}, \delta_1, \ldots, \delta_m\}$
- varying $\{\delta_1, \ldots, \delta_m\}$ at test time

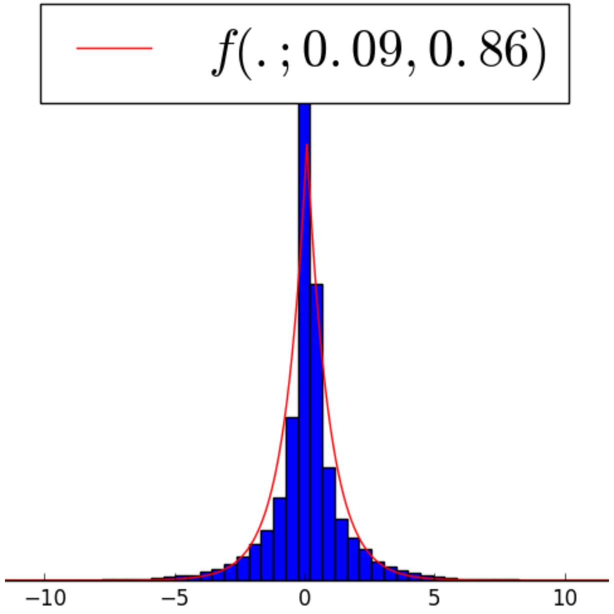
JPEG2000

III - Experiments

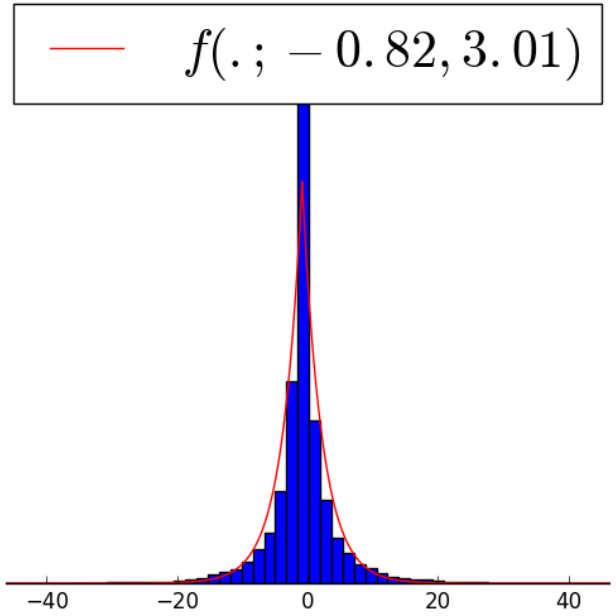


rate ≈ 0.23 bpp

IV - Interpretation



50^{th} feature map in $\boldsymbol{Y}$



125^{th} feature map in $\boldsymbol{Y}$

$$f(x;\mu,\lambda) = \frac{1}{2\lambda}\exp\left(\frac{|x-\mu|}{\lambda}\right)$$

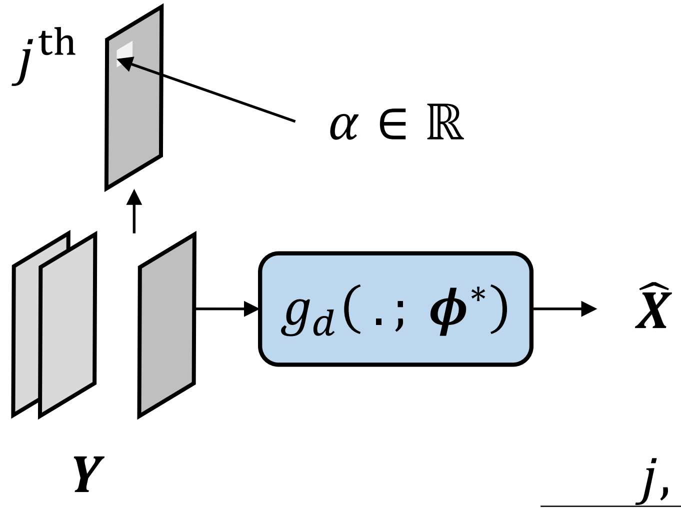
$\mu \in \mathbb{R}$

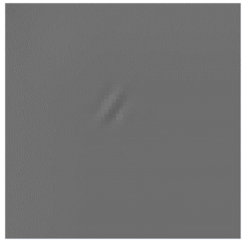
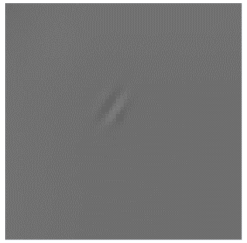
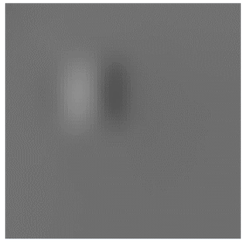
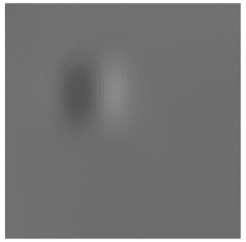
$\lambda \in {\mathbb{R}_+}^*$

Zero mean each feature map of $\boldsymbol{Y}$ ➡ **DCT-like distribution in each feature map** of $\boldsymbol{Y}$

IV - Interpretation

Was a DCT-like transform learned? No!



j, α	$50, 8.0$	$50, -8.0$	$125, 20.0$	$125, -20.0$
top-left patch of $\widehat{\boldsymbol{X}}$				

Thanks you for your attention!

For further details,

www.irisa.fr/temics/demos/visualization_ae/visualizationAE.htm