Posterior approximation with variational autoencoders applied to satellite image restoration

GDR-IASIS: Advances in learning-based image restoration

Maud Biquard[1,3]

Advisors: Thomas Oberlin[1], Marie Chabert[2], Florence Genin[3], Christophe Latry[3]

[maud.biquard@isae-supaero.fr]

9th december 2024

[1] ISAE-SUPAERO, [2] IRIT [3] CNES

[*Intro*] Problem and notations

Considered inverse problems

- Image restoration: deblurring and denoising
- Super-resolution: joint restoration and single image super-resolution

Forward model: $y = \mathcal{A}(x) + w$

Figure: Acquisition of the satellite and restored image

Bayesian Maximum A Posteriori Estimate (MAP):

$$\arg\max_x \underbrace{p(x|y)}_{\text{Posterior}} = \arg\min_x - \underbrace{\log p(y|x)}_{\text{Log-likelihood}} - \log \underbrace{p(x)}_{\text{Prior}} \tag{1}$$

[*Intro*] Content

[*Bckgrnd*] Generative neural networks

- To synthesize realistic data from a random variable
- Considered generative model:

$$p_\theta(x, z) = p_\theta(x|z) \underbrace{p_\theta(z)}_{\text{Latent prior}} \qquad (2)$$

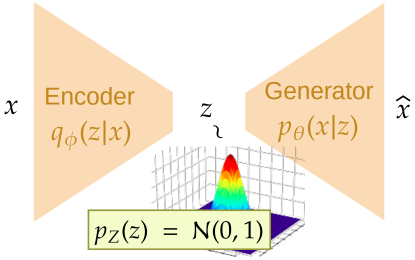


Figure: A variational autoencoder (VAE)

In the case of "classical" VAEs:

- Latent prior: $p_\theta(z)$ is often $\mathcal{N}(0, I)$
- Generative distribution (decoder): $p_\theta(x|z) = \mathcal{N}(D_\theta(z), \Sigma_\theta(z))$, often $\Sigma_\theta(z) = \gamma^2 I$
- Inference model (encoder): $q_\phi(z|x) \simeq p_\theta(z|x)$ also Gaussian

[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[Bckgrnd] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

Second step: Looking for the solution in the latent space of G.

Latent space Z

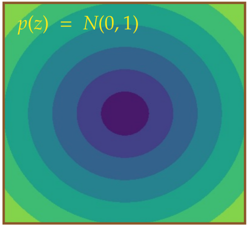


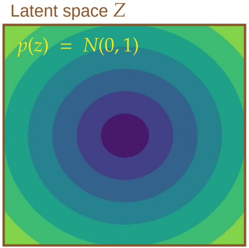
Figure: Restoration process of Bora's method[1]

[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

Second step: Looking for the solution in the latent space of G.



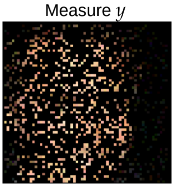


Figure: Restoration process of Bora's method[1]

[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

Second step: Looking for the solution in the latent space of G.

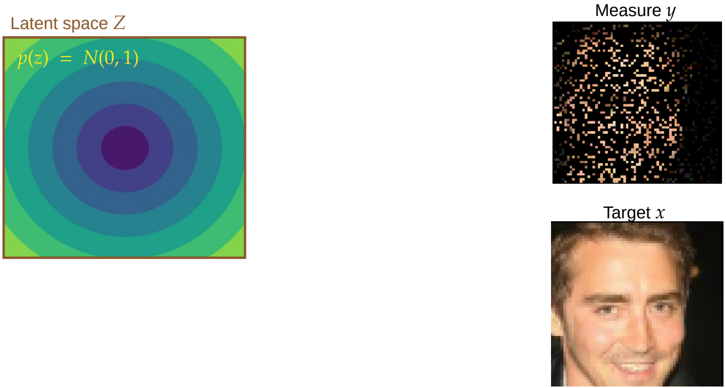


Figure: Restoration process of Bora's method[1]

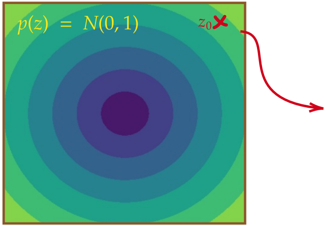
[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

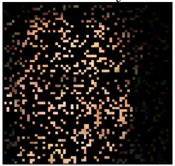
Second step: Looking for the solution in the latent space of G.

Latent space Z



$x_0 = G(z_0)$

Measure y



Target x

Optimization on z:

$z^* = \arg\min_z \; -\log p_{Y|Z}(y|z) - \log p_\theta(z)$

with $p_{Y|Z}(y|z) = p_{Y|X}(y|x = D_\theta(z))$

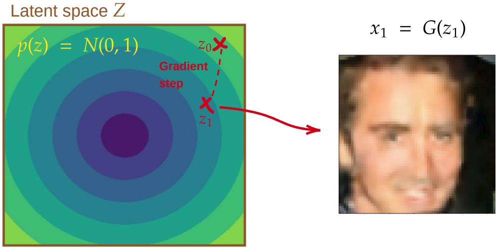
Figure: Restoration process of Bora's method[1]

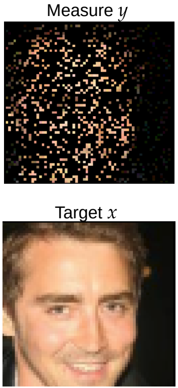
[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

Second step: Looking for the solution in the latent space of G.





Optimization on z:

$z^* = \arg\min_z -\log p_{Y|Z}(y|z) - \log p_\theta(z)$

with $p_{Y|Z}(y|z) = p_{Y|X}(y|x = D_\theta(z))$

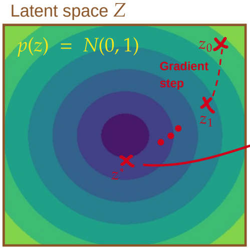
Figure: Restoration process of Bora's method[1]

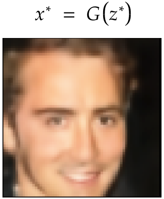
[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

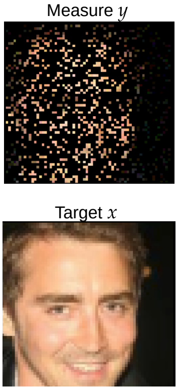
[*Bckgrnd*] Latent optimization methods[1]

First step: Train a generative model G on ideal images.

Second step: Looking for the solution in the latent space of G.







Optimization on z:

$z^* = \arg\min_z -\log p_{Y|Z}(y|z) - \log p_\theta(z)$

with $p_{Y|Z}(y|z) = p_{Y|X}(y|x = D_\theta(z))$

Figure: Restoration process of Bora's method[1]

[1] A.Bora, A.Jalal, E.Price, A.G.Dimakis, *Compressed Sensing using Generative Models*, ICML, 2017.

[VBLE] Variational Bayes Latent Estimation[1] (VBLE)

Approximation of the **latent posterior** $p(z|y)$ using Variational Inference:

- Parametric family for the approximated posterior:

$$\left\{ q_{\bar{z},a}(z) | \bar{z}, a \in \mathbb{R}^{C \times M \times N}, a > 0 \right\} \qquad (3)$$

with $q_{\bar{z},a}(z) = \prod_k \mathcal{N}(z_k; \bar{z}_k, a_k^2)$

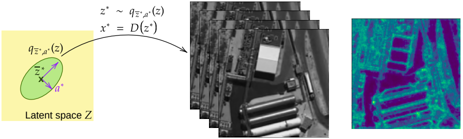


Figure: VBLE posterior sampling procedure

- Minimization of $KL(q_{\bar{z},a}(z)||p(z|y)) \Rightarrow$ maximization of the ELBO

$$\arg\max_{z,a} \mathbb{E}_{q_{\bar{z},a}(z)} \left[\log p(y|z) + \log p_\theta(z) - \log q_{\bar{z},a}(z) \right] \qquad (4)$$

[1] M. Biquard, M. Chabert, F. Genin, C. Latry, T. Oberlin, *Variational Bayes Image Restoration with compressive autoencoders*, 2024 (preprint)

[VBLExz] Limitation of latent posterior approximation

Three types of errors in latent optimization methods, due to:

- Inverse problem uncertainty
- Representation error ($x^* \notin \{D_\theta(z)\}_z$)
- Optimization

Hypothesis:

- Inverse problem uncertainty well modelled by VBLE
- But no modelling of the representation error

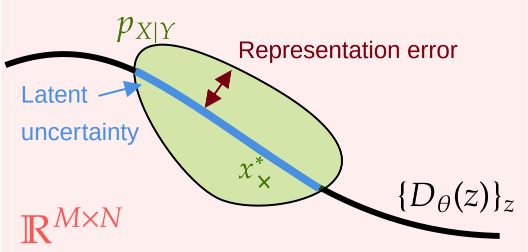


Figure: Representation error and latent uncertainty. x^*: inverse problem solution. $\{D_\theta(z)\}$ generator range.

[VBLExz] VBLE-xz algorithm[1]

Approximation of the **joint latent and image posterior** $p(x, z|y)$[2] of the inverse problem

- Supposing a generative decoder $p_\theta(x|z) = \mathcal{N}\left(D_\theta(z), \text{diag}(\sigma_\theta^2(z))\right)$
- Parametric family for the approximated posterior

$$\{q_b(x|z)q_{\bar{z},a}(z)\} \text{ with } q_b(x|z) = \prod_k \mathcal{N}(D_\theta(z)_k, (b\sigma_\theta(z))_k^2) \tag{5}$$

- Minimization of $KL(q_b(x|z)q_{\bar{z},a}(z)||p(x, z|y)) \Rightarrow$ maximization of the ELBO...

[1] Biquard, Chabert, Genin, Latry, Oberlin, *Deep priors for satellite image restoration*, 2024 (preprint)
[2] Gonzàlez, Almansa, Tan, *Solving inverse problems by joint posterior maximization with autoencoding prior*, SIAM, 2022

[*VBLExz*] Use of a variational compressive autoencoder (CAE)

- Can be seen as VAEs (with two latent variables) with the loss

$$\mathcal{L} = \underbrace{\alpha}_{\text{bitrate param}} \underbrace{||x - \hat{x}||_2^2}_{\text{Datafid}} + \underbrace{\text{Rate(z,h)}}_{\text{Latent constraint}}$$

Works well for image restoration as:

- The hyperprior = adaptive prior on the latent distribution $\rightarrow p(z) \propto \mathcal{N}(z; \mu^z, \sigma^z)$.
- Can adapt the bitrate (α) to the difficulty of the inverse problems

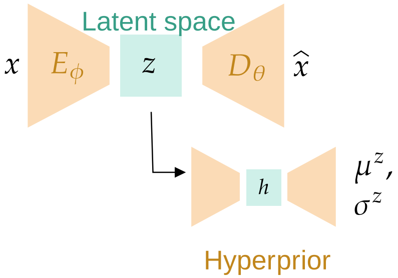


Figure: Compressive autoencoder with hyperprior[1]

[1] Ballé, Minnen, Singh, Hwang, Johnston, *Variational image compression with a scale hyperprior*, ICLR 2018

[*Results*] VBLE results on BSD[1]

Setup:

- Variational compressive autoencoder structure for VBLE
- BSD dataset
- Baselines: DPIR[2] (plug-and-play), DiffPIR[3] (diffusion based), PnP-ULA[4] (MCMC)

	#Pixels	**DPIR**	**DiffPIR**	**PnP-ULA**	**VBLE**
PSNR $\uparrow$	x	28.15	27.83	26.98	**28.30**
LPIPS $\downarrow$	x	0.2495	**0.2214**	0.2688	0.2430
Time	256^2	x	35min	1h23m	**27s**

Table: PSNR and LPIPS averaged on Gaussian deblur, SISR $\times 2$, SISR $\times 4$ problems. Time: computation time required for image restoration and sampling 100 posterior samples.

[1] Biquard et al., *Variational Bayes Image Restoration with compressive autoencoders*, 2024 (preprint)
[2] Zhang et al., *Plug-and-play image restoration with deep denoiser prior*, IEEE TPAMI, 2021
[3] Zhu et al., *Denoising Diffusion Models for Plug-and-Play Image Restoration*, CVPR, 2023
[4] Laumont et al., *Bayesian Imaging using plug play priors: when Langevin meets Tweedie*, SIAM, 2022

[*Results*] VBLE-xz results on satellite image restoration[1]

Baselines:

- Bay+IF: NL-Bayes + inverse filtering
- RDN[2] and SRResNet: direct inversion networks
- SatDPIR: proposed adapted plug-ang-play approach

		Bay+IF	**RDN**	**SRResNet**	**SatDPIR**	**VBLE-xz**
IR	PSNR ↑	40.56	48.08	48.44	**48.66**	48.19
	LPIPS ↓	0.0369	0.0145	0.0157	**0.0138**	0.0275
IR+SISR	PSNR ↑	33.22	36.66	36.28	**37.18**	36.63
	LPIPS ↓	0.2463	**0.1228**	0.1466	0.1658	0.1897

Table: Quantitative results on realistic Pléiades restoration. IR: Image Restoration. IR+SISR: combined IR and super-resolution.

[1] Biquard, Chabert, Genin, Latry, Oberlin, *Deep priors for satellite image restoration*, 2024 (preprint)
[2] Zhang et al, *Residual dense network for image super-resolution*, CVPR, 2018

[*Results*] VBLE-xz results on satellite image restoration (2)

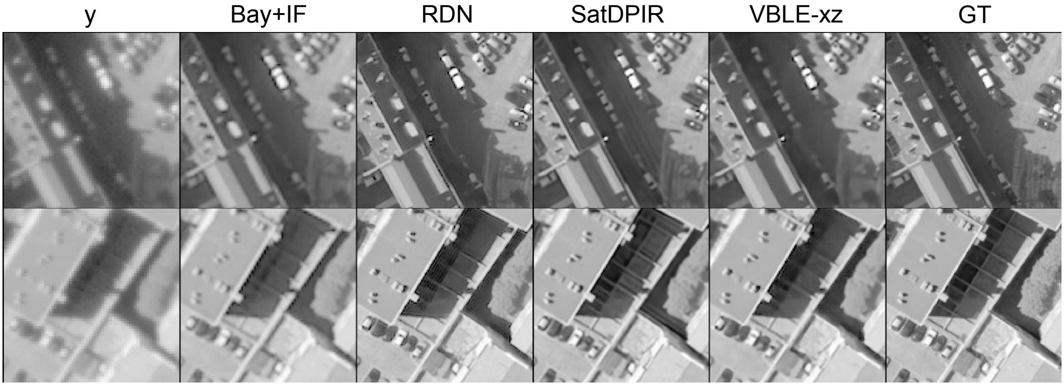


Figure: Qualitative results on IR+SISR methods. ©CNES2024

[*Results*] VBLE-xz results on satellite image restoration (3)

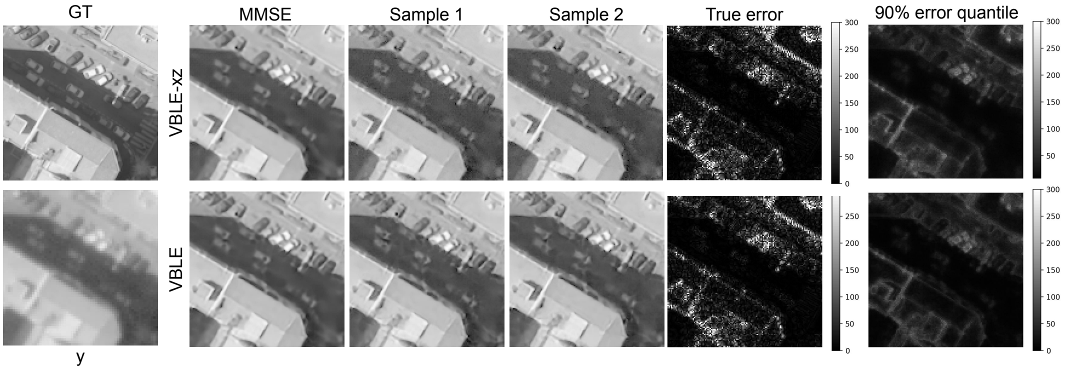


Figure: VBLE-xz VS VBLE sampling ability. ©CNES2024

[*Conclusion*] Conclusion and links

VBLE and VBLE-xz

- consist in approximating the inverse problem posterior in the latent (and image) space(s) of VAEs
- are nice alternatives for scalable posterior sampling

→ VBLE paper: *Variational Bayes image restoration with compressive autoencoders*
`https://arxiv.org/abs/2311.17744v3`

→ VBLE-xz paper: *Deep priors for satellite image restoration*
`http://arxiv.org/abs/2412.04130`

VBLE implementation on Github

[Appendice] Final VBLE algorithm

Reparameterization trick: $q_{\bar{z},a} = \bar{z} + a\epsilon$

$\rightarrow$ Stochastic Gradient Variational Bayes (SGVB) estimate can be derived[1]

Algorithm Variational Bayes Latent Estimation

Require: $\bar{z}_0 \in \mathbb{R}^{C\times M\times N}$, $a_0 \in \mathbb{R}^{C\times M\times N} = (1)_{i,j,l}$, $k=0$, $\eta > 0$
  while not *convergence* **do**
    $z_1 \sim q_{\bar{z}_k,a_k}(z_1), \ldots, z_n \sim q_{\bar{z}_k,a_k}(z_n)$
    $\begin{pmatrix} \bar{z}_{k+1} \\ a_{k+1} \end{pmatrix} = \begin{pmatrix} \bar{z}_k \\ a_k \end{pmatrix} - \eta \nabla_{\bar{z},a} \frac{1}{n} \sum_{i=1}^{n} \Big[-\log p_{Y|Z}(y|z_i) - \log p_\theta(z_i) + \log q_{\bar{z},a}(z_i) \big) \Big]$
    $k = k+1$
  end while
  return $(\bar{z}^*, a^*) = (\bar{z}_k, a_k)$

Final MMSE estimation:

$x^*_{MMSE-z} = D_\theta(\bar{z}^*)$ or $x^*_{MMSE-x} = \frac{1}{L}\sum_{i=1}^{L} D_\theta(z_i)$ with $z_i \sim q_{\bar{z}^*,a^*}(z_i)$.

[1] D.P. Kingma, M.Welling, *Auto-Encoding Variational Bayes*, ICLR 2014

[*Appendice*] VBLE-xz ELBO

VBLE-xz ELBO (general case):

$$\arg\max_{\bar{z},a,b} \mathbb{E}_{q_b(x|z)q_{\bar{z},a}(z)}\left[\log p_{Y|X}(y|x) + \log p_\theta(x|z) + \log p_\theta(z) - \log q_b(x|z) - \log q_{\bar{z},a}(z)\right]$$

VBLE-xz ELBO (VAE with Gaussian generative and inference models):

$$\arg\max_{\bar{z},a,b} \mathbb{E}_{q_b(x|z)q_{\bar{z},a}(z)}\left[\log p_{Y|X}(y|x)\right] - \mathrm{KL}\left[q_{\bar{z},a}(z)||p_\theta(z)\right] - \mathbb{E}_{q_{\bar{z},a}(z)}\left(\mathrm{KL}\left[q_b(x|z)||p_\theta(x|z)\right]\right)$$

$$= \arg\min_{\bar{z},a,b} \mathbb{E}_{q_b(x|z)q_{\bar{z},a}(z)}\left[-\log p_{Y|X}(y|x)\right] + \sum_k \left[\log a_k - \frac{1}{2}(a_k^2 + \bar{z}^2)\right] + \sum_i \log b_i - \frac{1}{2}b_i^2.$$

[*Appendice*] Training of a VAE with decoder variance

Step 1: training a VAE with a fixed decoder variance, that is $p_{\theta_1}(x|z) = \mathcal{N}(D_{\theta_1}(z), \gamma^2 I)$

$$\mathcal{L}_1(\theta_1, \phi; x) = \mathbb{E}_{q_\phi(z|x)}\left[\frac{1}{2\gamma^2}||x - D_{\theta_1}(z)||^2\right] + \mathrm{KL}\left[q_\phi(z|x)||p_{\theta_1}(z)\right]$$

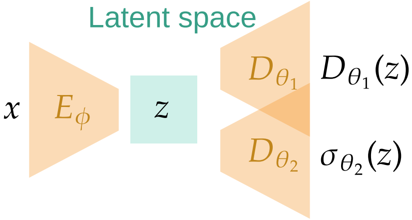


Figure: VAE predicting the decoder variance.

Step 2: training the decoder variance, by minimizing the negative log-likelihood

$$\mathcal{L}_1(\theta_2; x) = \mathbb{E}_{q_\phi(z|x)}\left[\sum_k \left[\frac{(x_k - D_{\theta_1,k}(z))^2}{2\sigma^2_{\theta_2,k}(z)}\right] + \log \sigma_{\theta_2,k}(z)\right]$$

[*Appendice*] Use of a compressive autoencoder (CAE) - Details

CAEs can generally be expressed as VAEs.

- Uniform encoder posterior distribution

$$q_\phi(z, h|x) = q_\phi(z|x, h)q_\phi(h|x)$$

$$\text{with } q_\phi(z|x, h) = \prod_k \mathcal{U}([z_k - \frac{1}{2}, z_k + \frac{1}{2}])$$

- Hyperprior $\simeq$ 2 latent variable VAE with

$$p_\theta(z, h) = p_\theta(z|h)p_\theta(h)$$

$$\text{and } p_\theta(z|h) = \prod_k \left[\mathcal{N}(\mu^z, \sigma^z) * \mathcal{U}[-\frac{1}{2}, \frac{1}{2}])\right](z_k)$$

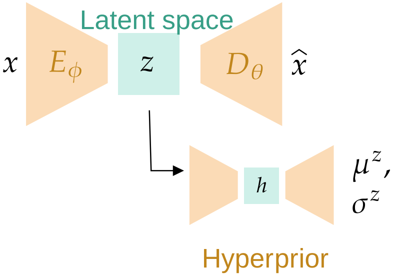


Figure: Compressive autoencoder with hyperprior[1]

[1] J. Ballé, D. Minnen, S. Singh, S.J. Hwang, N. Johnston, *Variationnal image compression with a scale hyperprior*, ICLR 2018

[*Appendice*] CAE expressed in the VAE framework

CAE generative model

$$p_\theta(x,z,h) = p_\theta(x|z)p_\theta(z|h)p_\theta(h) \text{ with } p_\theta(x|z) = \prod_k \mathcal{N}\big(x_k;\ D_\theta(z)_k, \frac{1}{2\alpha\log(2)}\big),$$

$$p_\theta(z|h) = \prod_k \Big[\mathcal{N}(z_k;\ \mu_k^z, (\sigma_k^z)^2) * U\big(z_k;\ [-\frac{1}{2},\frac{1}{2}]\big)\Big] \text{ and } p_\theta(h) = \prod_k p_\psi(h_k)$$

CAE inference model

$$q_\phi(z,h|x) = q_\phi(z|x,h)q_\phi(h|x) \text{ with } q_\phi(z|x,h) = \prod_k \mathcal{U}(z_k;\ [\bar{z}_k - \frac{1}{2}, \bar{z}_k + \frac{1}{2}])$$

$$\text{and } q_\phi(h|x) = \prod_k \mathcal{U}(h_k;\ [\bar{h}_k - \frac{1}{2}, \bar{h}_k + \frac{1}{2}]). \tag{6}$$

Then the ELBO (1st line) corresponds to the rate distortion loss (2nd line)

$$\begin{aligned}\mathcal{L}(x) &= \mathbb{E}_{q_\phi(z,h|x)}\big[\log q_\phi(z,h|x) - \log p_\theta(x|z,h) - \log p_\theta(z,h)\big] \qquad (7)\\ &\propto 0 + \log(2)(\alpha \text{Distortion}(x,z) + \text{ Rate}(z,h)).\end{aligned}$$