Inverse Problem Regularization with a Variational Autoencoder Prior

CANUM 2024

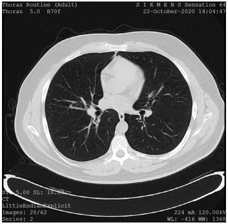
Jean Prost[1], Antoine Houdard[2], Andrés Almansa[3], Nicolas Papadakis[4]

[1] CRIStAL Lille
[2] Ubisoft Bordeaux
[3] MAP5, Université Paris Cité
[4] Institut de Mathématiques de Bordeaux

May 27, 2024

CRIStAL
Centre de Recherche en Informatique, Signal et Automatique de Lille

Image restoration



Medical image

Astrophysic

Photography

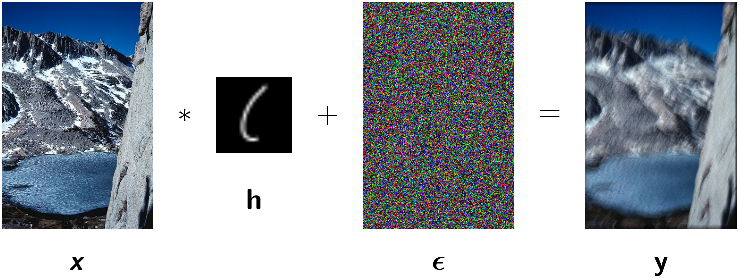
- Sensor limitations
- Incomplete measurements
- External factors (low-light, movement)

→ Need to develop image restoration algorithms

Linear inverse problem

Restoring an image amounts to solve an image inverse problem:

$$\boldsymbol{y} = H\boldsymbol{x} + \epsilon, \qquad \epsilon \sim \mathcal{N}(\epsilon; 0, \sigma^2 I) \tag{1}$$



Goal: find $\boldsymbol{x}$!

Variational problem

We can recover a clean image $\boldsymbol{x}$ by solving an optimization problem:

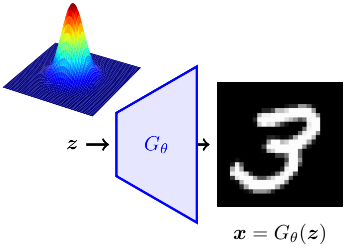
$$\hat{\boldsymbol{x}}_{map} = \arg\min_{\boldsymbol{x}} \underbrace{\frac{1}{2\sigma^2}||H\boldsymbol{x} - \boldsymbol{y}||^2}_{\text{likelihood}} + \underbrace{g(\boldsymbol{x})}_{\text{prior}} \tag{2}$$

Under a Bayesian perspective $g(\boldsymbol{x}) = -\log p(\boldsymbol{x})$

Classical prior:

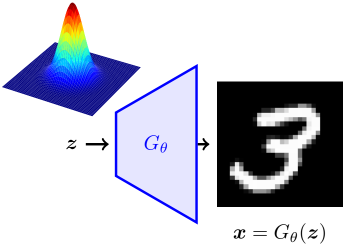
- Sparsity: $g(\boldsymbol{x}) = \|W\boldsymbol{x}\|_1$
- Total variation: $g(\boldsymbol{x}) = \|\nabla\boldsymbol{x}\|_2$
- ...

Deep generative model:



Variational autoencoder

Generative model:

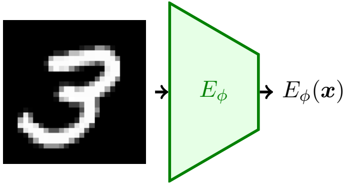


- Prior:
$$p_\theta(\boldsymbol{z}) = \mathcal{N}(0, \mathrm{Id})$$
- Decoder:
$$p_\theta(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}\left(G_\theta(\boldsymbol{z}), \gamma^2 \, \mathrm{Id}\right)$$

Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". *arXiv preprint arXiv:1312.6114* (2013)

Variational autoencoder

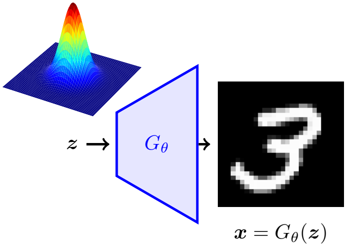
Encoder:



- Inference model:

$$q_\phi(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}\left(E_\phi(\boldsymbol{x}), \Sigma_\phi(\boldsymbol{x})\right)$$

Generative model:



- Prior:

$$p_\theta\left(\boldsymbol{z}\right) = \mathcal{N}\left(0, \mathrm{Id}\right)$$

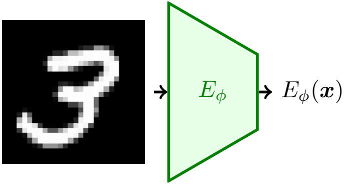
- Decoder:

$$p_\theta\left(\boldsymbol{x}|\boldsymbol{z}\right) = \mathcal{N}\left(G_\theta(\boldsymbol{z}), \gamma^2\,\mathrm{Id}\right)$$

Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. *arXiv preprint arXiv:1312.6114* (2013)

Variational autoencoder

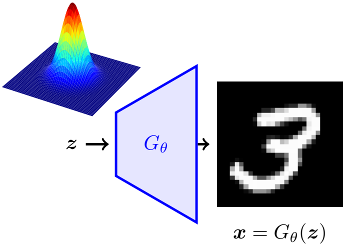
Encoder:



- Inference model:

$$q_\phi(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}\left(E_\phi(\boldsymbol{x}), \Sigma_\phi(\boldsymbol{x})\right)$$

Generative model:



- Prior:

$$p_\theta\left(\boldsymbol{z}\right) = \mathcal{N}\left(0, \mathrm{Id}\right)$$

- Decoder:

$$p_\theta\left(\boldsymbol{x}|\boldsymbol{z}\right) = \mathcal{N}\left(G_\theta(\boldsymbol{z}), \gamma^2\,\mathrm{Id}\right)$$

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

Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. *arXiv preprint arXiv:1312.6114* (2013)

Joint posterior maximization with autoencoding prior

Goal: use the probabilistic model learned by the VAE as a prior:

$$p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z})p_\theta(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \tag{3}$$

- Compute an "augmented" MAP estimator:

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \underbrace{\frac{1}{2\sigma^2}\left|\left|H\boldsymbol{x} - \boldsymbol{y}\right|\right|^2}_{\text{data fidelity}} + \underbrace{\frac{1}{2\gamma^2}\left|\left|\boldsymbol{x} - G_\theta\left(\boldsymbol{z}\right)\right|\right|^2}_{x-z\text{ coupling}} + \underbrace{\frac{1}{2}\|\boldsymbol{z}\|^2}_{\text{latent prior}}$$

Mario González, Andrés Almansa, and Pauline Tan. "Solving inverse problems by joint posterior maximization with autoencoding prior". *SIAM Journal on Imaging Sciences* 2 (2022)

Joint posterior maximization with autoencoding prior

Goal: use the probabilistic model learned by the VAE as a prior:

$$p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z}) p_\theta(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \tag{3}$$

- Compute an "augmented" MAP estimator:

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \underbrace{\frac{1}{2\sigma^2} ||H\boldsymbol{x} - \boldsymbol{y}||^2}_{\text{data fidelity}} + \underbrace{\frac{1}{2\gamma^2} ||\boldsymbol{x} - G_\theta(\boldsymbol{z})||^2}_{x-z \text{ coupling}} + \underbrace{\frac{1}{2}\|\boldsymbol{z}\|^2}_{\text{latent prior}}$$

- Alternate optimization (JPMAP):

$$\boldsymbol{z}^{n+1} = \arg\min_{\boldsymbol{z}} \frac{1}{2\gamma^2}\|\boldsymbol{x}^n - G_\theta(\boldsymbol{z})\|^2 + \frac{1}{2}\|\boldsymbol{z}\|^2 \quad ?$$

$$\boldsymbol{x}^{n+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2\sigma^2}\|H\boldsymbol{x} - \boldsymbol{y}\|^2 + \frac{1}{2\gamma^2}\|\boldsymbol{x} - G_\theta\left(\boldsymbol{z}^{n+1}\right)\|^2 \quad \checkmark$$

Mario González, Andrés Almansa, and Pauline Tan. "Solving inverse problems by joint posterior maximization with autoencoding prior". *SIAM Journal on Imaging Sciences* 2 (2022)

Joint posterior maximization with autoencoding prior

Goal: use the probabilistic model learned by the VAE as a prior:

$$p_\theta(\boldsymbol{x}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z}) p_\theta(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \tag{3}$$

- Compute an "augmented" MAP estimator:

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \underbrace{\frac{1}{2\sigma^2} ||H\boldsymbol{x} - \boldsymbol{y}||^2}_{\text{data fidelity}} + \underbrace{\frac{1}{2\gamma^2} ||\boldsymbol{x} - G_\theta(\boldsymbol{z})||^2}_{x-z \text{ coupling}} + \underbrace{\frac{1}{2}\|\boldsymbol{z}\|^2}_{\text{latent prior}}$$

- Alternate optimization (JPMAP):

$$\boldsymbol{z}^{n+1} = \arg\min_{\boldsymbol{z}} \frac{1}{2\gamma^2}\|\boldsymbol{x}^n - G_\theta(\boldsymbol{z})\|^2 + \frac{1}{2}\|\boldsymbol{z}\|^2 \approx E_\phi(\boldsymbol{x}^{n+1})$$

$$\boldsymbol{x}^{n+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2\sigma^2}\|H\boldsymbol{x} - \boldsymbol{y}\|^2 + \frac{1}{2\gamma^2}\|\boldsymbol{x} - G_\theta\left(\boldsymbol{z}^{n+1}\right)\|^2 \quad \checkmark$$

Mario González, Andrés Almansa, and Pauline Tan. "Solving inverse problems by joint posterior maximization with autoencoding prior". *SIAM Journal on Imaging Sciences* 2 (2022)

Convergence

Alternate convex search[1]

JPMAP converges if:

- The encoder is "perfect":

$$q_\phi(\boldsymbol{z}|\boldsymbol{x}) = p_\theta(\boldsymbol{z}|\boldsymbol{x})$$

- $G_\theta(\boldsymbol{z})$ is continuously differentiable

Fixed point convergence[2]

If the JPMAP iteration:

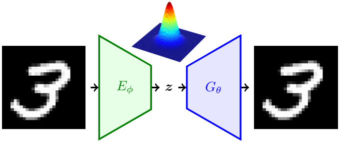
$$\boldsymbol{x}^{n+1} = \mathrm{prox}_{\gamma^2 f}\left(G_\theta\left(E_\phi(\boldsymbol{x}^n)\right)\right) \quad \text{with} \quad f(\boldsymbol{x}) := \frac{1}{2\sigma^2}\left\|H\boldsymbol{x} - \boldsymbol{y}\right\|^2$$

is contractive, JPMAP converges to a fixed point.

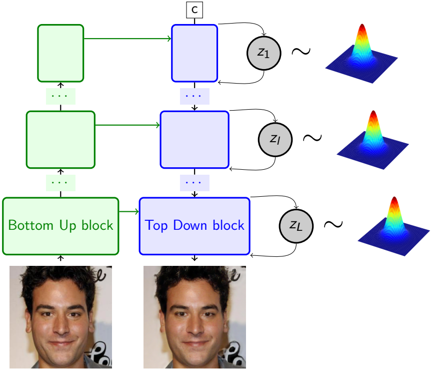
[1] Mario González, Andrés Almansa, and Pauline Tan. "Solving inverse problems by joint posterior maximization with autoencoding prior". *SIAM Journal on Imaging Sciences* 2 (2022).

[2] Jean Prost et al. "Inverse problem regularization with hierarchical variational autoencoders". *arXiv preprint arXiv:2303.11217* (2023).

Scaling JPMAP to Hierarchical VAEs



VAE



Hierarchical VAE

Rewon Child. "Very deep vaes generalize autoregressive models and can outperform them on images". *arXiv preprint arXiv:2011.10650* (2020)

Challenges

JPMAP iteration:

$$x^{n+1} = \mathrm{prox}_{\gamma^2 f}\left(G_\theta\left(E_\phi(x^n)\right)\right) \quad \text{with} \quad f(x) := \frac{1}{2\sigma^2}\left\|Hx - y\right\|^2$$

For a high capacity HVAE:

$$G_\theta\left(E_\phi(x)\right) \approx x$$

blurry JPMAP 26.1dB GT

→ no regularization :(

Temperature scaling

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2\sigma^2} ||H\boldsymbol{x} - \boldsymbol{y}||^2 + \frac{1}{2\gamma^2} ||\boldsymbol{x} - G_\theta(\boldsymbol{z})||^2 - \frac{1}{\tau^2} \log p_\theta(\boldsymbol{z})$$

The $\boldsymbol{z} - \textit{step}$ becomes:

$$\boldsymbol{z}^{n+1} = \arg\max_{\boldsymbol{z}} \log \underbrace{q_\phi(\boldsymbol{z}|\boldsymbol{x}^n)}_{\text{encoder}} + \left(\frac{1}{\tau^2} - 1\right) \log \underbrace{p_\theta(\boldsymbol{z})}_{\text{prior}} := E_{\phi,\boldsymbol{\tau}}(\boldsymbol{x}^n)$$

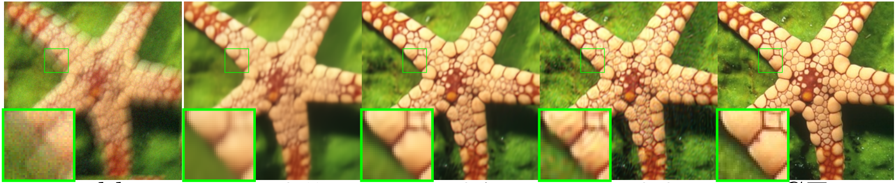


Figure: Motion deblurring, $\sigma = 7.65$

Face image restoration

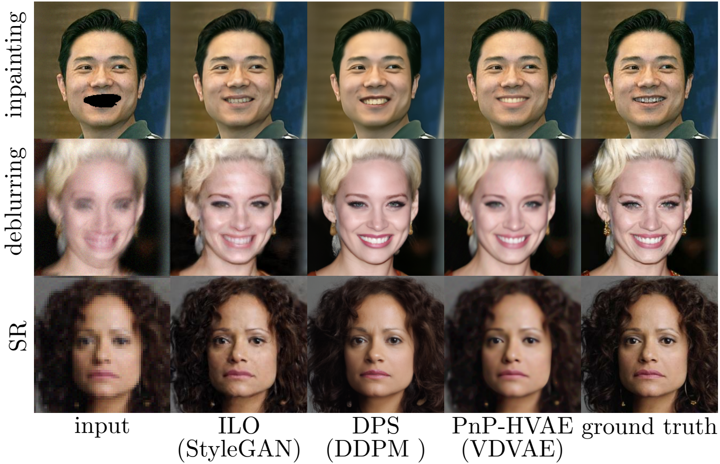
- Experiments on CelebA dataset
- Test on image inpainting, deblurring, super-resolution
- We used a pretrained VDVAE from[3]
- Compared method:
  - Diffusion posterior sampling (DPS)[4] (DDPM)
  - ILO[5] (StyleGAN inversion)
  - Optimizing $J(\boldsymbol{x}, \boldsymbol{z})$ with Adam

[3] Rewon Child. "Very deep vaes generalize autoregressive models and can outperform them on images". *arXiv preprint arXiv:2011.10650* (2020).

[4] Hyungjin Chung et al. "Diffusion Posterior Sampling for General Noisy Inverse Problems". Sept. 2023.

[5] Giannis Daras et al. "Intermediate Layer Optimization for Inverse Problems using Deep Generative Models". PMLR. 2021.

Face images restoration: results



- ILO and DPS generate sharp images
- but PnP-HVAE is closer to the ground truth

Face images restoration: results

		PSNR↑	SSIM↑	LPIPS↓	time (s)
SR $\times 4$ $\sigma = 3$	Adam	28.56	0.75	0.38	<u>26</u>
	ILO	<u>28.80</u>	<u>0.78</u>	**0.17**	34
	DPS	27.53	0.76	<u>0.21</u>	153
	PnP-HVAE	**29.32**	**0.82**	0.28	**15**
Deblurring (motion) $\sigma = 8$	Adam	24.37	0.66	0.37	<u>12</u>
	ILO	<u>29.01</u>	<u>0.80</u>	<u>0.20</u>	34
	DPS	28.70	<u>0.80</u>	0.23	142
	PnP-HVAE	**30.40**	**0.84**	**0.16**	**10**
Deblurring (Gaussian) $\sigma = 8$	Adam	28.59	0.78	<u>0.23</u>	<u>12</u>
	ILO	29.12	0.79	**0.17**	34
	DPS	<u>29.14</u>	<u>0.81</u>	0.23	142
	PnP-HVAE	**30.81**	**0.86**	0.24	**10**

- Faster than concurent generative regularization methods (no backprop)
- PnP-HVAE provides the best results in PSNR and SSIM (more consistent)

Image deblurring

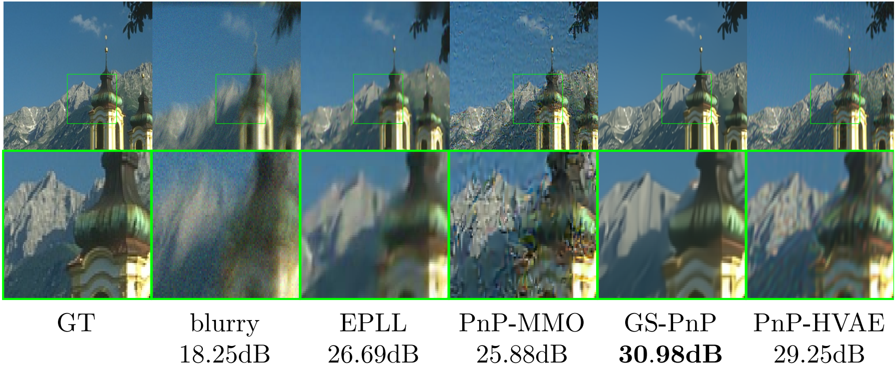
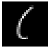
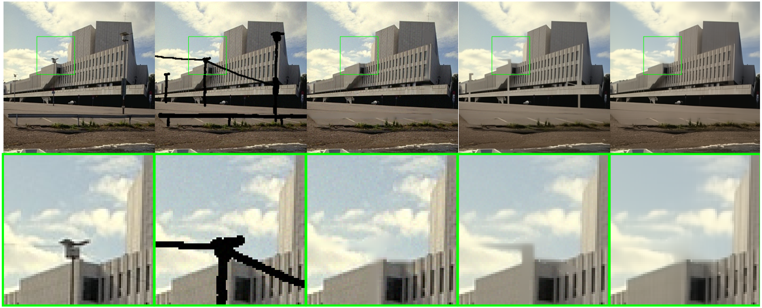


Figure: $\sigma = 7.65$, Motion blur 

- Fully convolutional HVAE trained on natural images patches
- Still outperformed by SOTA denoising PnP methods
- Is more robust to high noise-level than EPLL or PnP-MMO

Image inpainting



GT Masked EPLL GS-PnP PnP-HVAE

- PnP-HVAE is better to recreate missing information

Conclusion

We propose a method to use an HVAE model as a prior to solve inverse problems:

- Fixed point convergence under reasonable assumptions
- Memory efficient (no backpropagation)
- Results on par with other generative regularization methods
- Still room for improvement on natural images?

Temperature scaling

We control the regularization via a temperature parameter τ:

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x},\boldsymbol{z}} -\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p_\theta(\boldsymbol{x}|\boldsymbol{z}) - \frac{1}{\tau^2} \log p_\theta(\boldsymbol{z}) \tag{4}$$

This amounts to define a new prior:

$$p_{\theta,\boldsymbol{\tau}}(\boldsymbol{z}) \propto p_\theta(\boldsymbol{z})^{\frac{1}{\tau^2}} \tag{5}$$

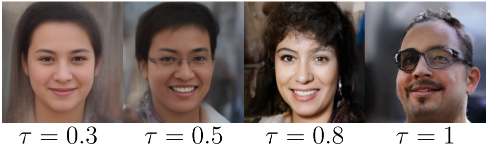
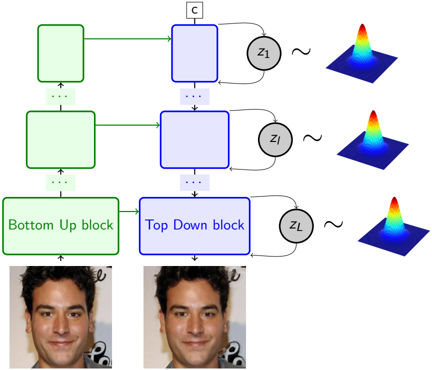


Figure: samples $G_\theta(\boldsymbol{z})$ with $\boldsymbol{z} \sim p_{\theta,\boldsymbol{\tau}}(\boldsymbol{z})$

Rewon Child. "Very deep vaes generalize autoregressive models and can outperform them on images". *arXiv preprint arXiv:2011.10650* (2020)

Hierarchical Variational autoencoder



- Hierarchical generative model

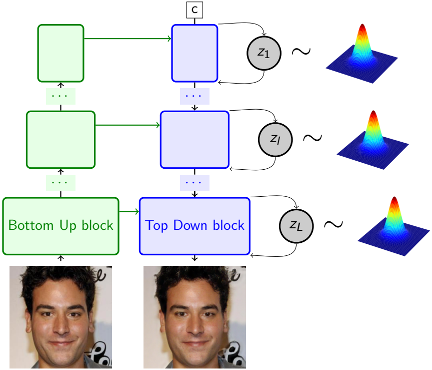
$$p_\theta(\mathbf{z}) := p_\theta(\mathbf{z}_1) \prod_{l=2}^{L} p_\theta(\mathbf{z}_l | \mathbf{z}_{<l})$$

- Hierarchical encoder

$$q_\phi(\mathbf{z}|\mathbf{x}) = q_\phi(\mathbf{z}_1|\mathbf{x}) \prod_{l=2}^{L} q_\phi(\mathbf{z}_l | \mathbf{z}_{<l}, \mathbf{x})$$

Rewon Child. "Very deep vaes generalize autoregressive models and can outperform them on images". *arXiv preprint arXiv:2011.10650* (2020)

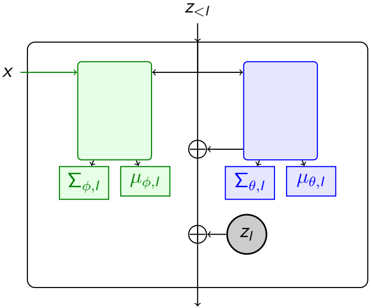
Hierarchical Variational autoencoder



- Gaussian marginals

$$p_\theta(\mathbf{z}_l|\mathbf{z}_{<l}) = \mathcal{N}(\mathbf{z}_l; \mu_{\theta_l}, \Sigma_{\theta,l})$$

$$q_\phi(\mathbf{z}_l|\mathbf{z}_{<l}, \mathbf{x}) = \mathcal{N}(\mathbf{z}_l; \mu_{\phi,l}, \Sigma_{\phi,l})$$



Rewon Child. "Very deep vaes generalize autoregressive models and can outperform them on images". *arXiv preprint arXiv:2011.10650* (2020)

Sequential encoding

How to solve the $\boldsymbol{z} - \textit{step}$ with a HVAE?

- Optimize sequentially each latent group
- For each level l interpolate between the encoder and the prior

---

Algorithm Sequential z-step

$\hat{\boldsymbol{z}}_1 \leftarrow \arg\max_{\boldsymbol{z}_1} \log q_\phi\left(\boldsymbol{z}_1 | \boldsymbol{x}^{(n)}\right) + \left(\frac{1}{\tau^2} - 1\right) \log p_\theta(\boldsymbol{z}_1)$

for $2 \leq l \leq L$ **do**

$\quad \hat{\boldsymbol{z}}_l = \arg\max_{\boldsymbol{z}_l} \log q_\phi\left(\boldsymbol{z}_l | \boldsymbol{z}_{<l}, \boldsymbol{x}^{(n)}\right) + \left(\frac{1}{\tau^2} - 1\right) \log p_\theta(\boldsymbol{z}_l | \boldsymbol{z}_{<l})$

end for

return $E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^{(n)}\right) = (\hat{\boldsymbol{z}}_1, \hat{\boldsymbol{z}}_1, \cdots, \hat{\boldsymbol{z}}_L)$

Final algorithm

In summary, our alternate optimization algorithm is:

$$\boldsymbol{z}^{(n+1)} = E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^{(n)}\right) \tag{6}$$

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} \underbrace{\frac{1}{2\sigma^2}\left\|H\boldsymbol{x}-\boldsymbol{y}\right\|^2}_{f(\boldsymbol{x})} + \frac{1}{2\gamma^2}\left\|\boldsymbol{x} - G_\theta\left(\boldsymbol{z}^{(n+1)}\right)\right\|^2 \tag{7}$$

or:

$$\boldsymbol{x}^{n+1} = \mathrm{prox}_{\gamma^2 f}\left(G_\theta\left(E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^n\right)\right)\right) \tag{8}$$

Like a Plug-and-play algorithm with $G_\theta\left(E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}\right)\right)$ as a "denoiser".

Fixed-point convergence

An iteration writes:

$$\begin{aligned} \boldsymbol{x}^{n+1} &= \mathrm{prox}_{\gamma^2 f}\left(G_\theta\left(E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^n\right)\right)\right) \\ &= T(\boldsymbol{x}^n) \end{aligned}$$

Let:

- λ_{min} the smallest value of $H^t H$
- L_τ the Lipschitz constant of $\boldsymbol{x} \to G_\theta\left(E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^n\right)\right)$

We have:

$$||T(\boldsymbol{u}) - T(\boldsymbol{v})|| \leq \frac{\sigma^2}{\gamma^2 \lambda_{\text{min}} + \sigma^2} L_\tau ||\boldsymbol{u} - \boldsymbol{v}||. \quad (9)$$

If $L_\tau < \frac{\gamma^2}{\sigma^2}\lambda_{min} + 1$, the iterations converge to a fixed-point $\boldsymbol{x}^\star$:

Properties of the fixed-point

- The fixed-point $\boldsymbol{x}^\star$ verifies:

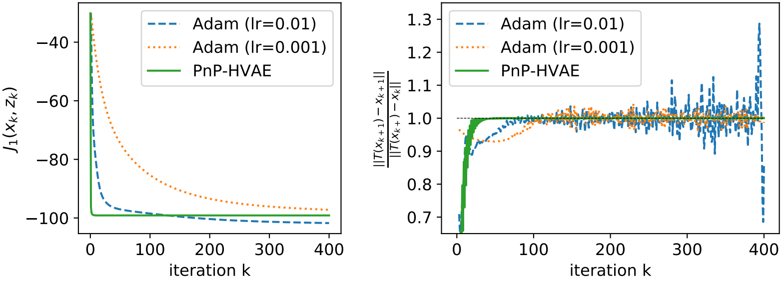
$$\nabla f(\boldsymbol{x}^\star) = \frac{1}{\gamma^2}\left(G_\theta\left(E_{\phi,\boldsymbol{\tau}}\left(\boldsymbol{x}^\star\right)\right) - \boldsymbol{x}^\star\right) \tag{10}$$

- We don't need a perfect encoder to guarantee the convergence!

- Under additional assumptions on the encoder, we have:

$$\nabla_{\boldsymbol{x}}\left(-\log p(\boldsymbol{y}|\boldsymbol{x}^\star) - \log p_{\theta,\boldsymbol{\tau}}(\boldsymbol{x}^\star)\right) = 0 \tag{11}$$

Numerical convergence



Comparison with a naive optimization with Adam of the same criterion:

- PnP-HVAE converges faster
- Empirically, $\frac{\left\| T(\boldsymbol{x}^{k+1}) - T(\boldsymbol{x}^{k}) \right\|}{\left\| \boldsymbol{x}^{k+1} - \boldsymbol{x}^{k} \right\|} \leq 1$ along the iterations

Temperature effect

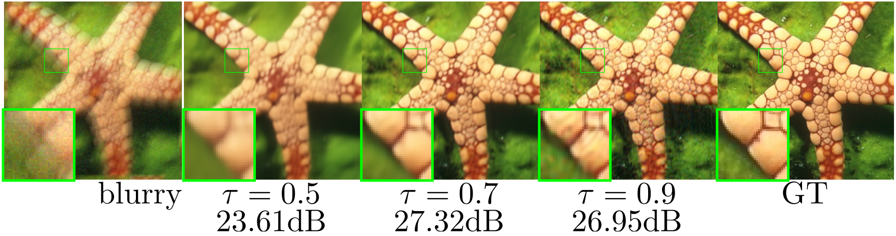


Figure: deblurring, $\sigma = 7.65$

- Low-temperature: Over-smoothing
- $\tau \to 1$: No regularization $\left(G_\theta\left(E_{\phi,\boldsymbol{\tau}}(x)\right) \approx \boldsymbol{x}\right)$