Deep regularization of inverse problems for satellite imaging

Gdr-mia workshop: Mathematical Models for Plug-and-play Image Restoration

Maud Biquard[1]

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

[maud.biquard@isae-supaero.fr]

December 08, 2022

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

[*Introduction*] Satellite imaging and image restoration

Images are degraded during their acquisition because of

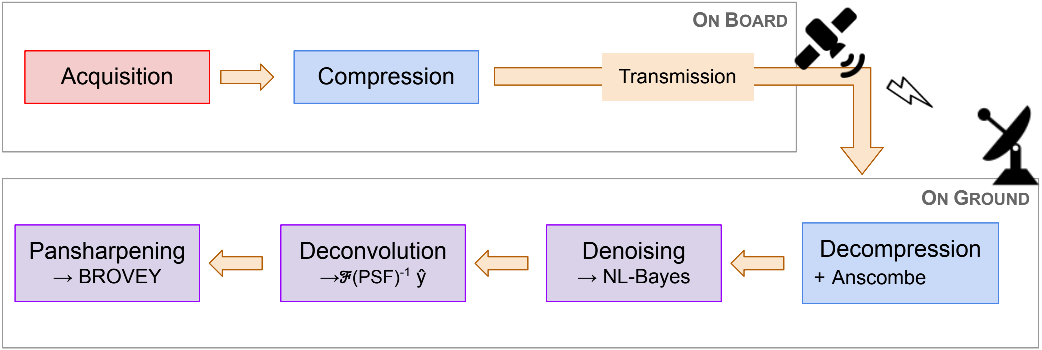
- sensors and transmission (noise, blur, compression artefacts)
- the acquisition conditions (movement, atmospheric perturbations)

Figure: Acquisition of the satellite and restored image

Other challenges to improve image quality $\rightarrow$ super-resolution, pan sharpening

$\Rightarrow$ Images need to be *restored* for further uses.

[*Introduction*] CNES image restoration pipeline[1]



[1] J.M. Delvit, C. Thiebaut, C. Latry, G. Blanchet, R. Camarero, *A pipeline to improve compressed image quality*, ICSO 2018

[*Introduction*] Data-based method requirements

Adaptive to the inverse problem $\Rightarrow$ data-driven model independent of the degradation

- Variety of on-board sensors
- Drift of sensors' characteristics over time

Explicitly using the forward model

- Forward model precisely known
- More interpretability

Easily tunable regularization parameter

- Rather easy inverse problems
- Complex data with highly varying statistics

[*Problem*] Inverse problem

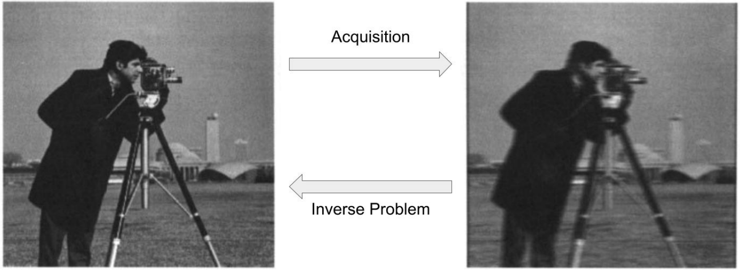


Image $x \in X$ Measurement $y \in Y$

Degradation model

$$y = \mathcal{A}(x) + n$$

Example of problems

$$y = x + n \qquad (denoising)$$

$$y = h * x + n \qquad (deblurring)$$

[*Problem*] Neural networks based image restoration methods

Learning to inverse:

- Supervised learning of a function $f : Y \to X$ from samples (x_i, y_i)
- Specific to the problem $\Rightarrow$ a network for *each* degradation

Learning only the regularisation from (x_i):

- No need of simulated images (y_i)
- Generic method $\Rightarrow$ adaptive to the degradation

[*Problem*] Neural networks based image restoration methods

Learning to inverse:

- Supervised learning of a function $f : Y \to X$ from samples (x_i, y_i)
- Specific to the problem $\Rightarrow$ a network for *each* degradation

Learning only the regularisation from (x_i):

- No need of simulated images (y_i)
- Generic method $\Rightarrow$ adaptive to the degradation

Using denoisers

Using generative models

[*Problem*] Generative neural networks

- To synthesize realistic data
- Different types : GAN, VAE, Normalizing Flows, Diffusion Models
- Function G that maps $z \sim p_Z$ to $x = G(z)$ from the image distribution p_X

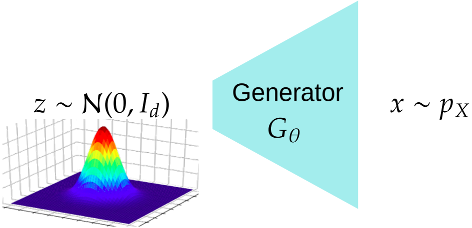


Figure: A generative model

[*Problem*] Compressed sensing using Generative Models[1]

Looking for the solution in the latent space of a generator G.

$$\hat{x} = G(\hat{z}) \text{ where } \hat{z} = \underset{z}{\operatorname{argmin}} ||AG(z) - y||^2 + \lambda ||z||^2 \tag{1}$$

$\sim$ MAP estimation in the latent space

Impressive results, but drawbacks:

- The solution $\hat{x} = G(\hat{z})$ lies in the range of the generator

$\Rightarrow$ Need for an excellent generative model

$\Rightarrow$ Even with it, not robust to out-of-distribution samples

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

[*Contributions*] Analysis formulation

Initial problem looks like a synthesis formulation:

$\hat{z} = \text{argmin}_z \, ||AG(z) - y||^2 + \lambda||z||^2$

$$\downarrow$$

An analysis formulation would be:

$\hat{x} = \text{argmin}_x \, ||Ax - y||^2 + \lambda||\mu_\phi(x)||_2^2$

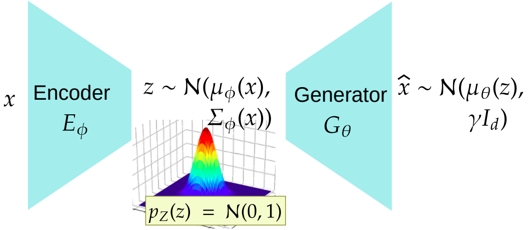


Figure: A variational autoencoder

- Any $\hat{x} \in \mathbb{R}^{M \times N}$ potentially reachable
- is not working with VAE
- $\rightarrow$ found solution has a high reconstruction loss

[1] M. Duff N.D. Campbell, M.J. Ehrhardt, *Regularising Inverse Problems with Generative Machine Learning Models*, arXiv preprint, 2021.
[2] M. González, A. Almansa, P. Tan, *Solving inverse problems by joint posterior maximization with autoencoding prior*, SIAM, 2022.
[3] S.A. Bigdeli, M. Zwicker, P. Tan, *Image restoration using autoencoding priors*, VISIGRAPP, 2018.

[*Contributions*] Analysis formulation

Initial problem looks like a synthesis formulation:

$\hat{z} = \text{argmin}_z \, ||AG(z) - y||^2 + \lambda ||z||^2$

$$\downarrow$$

An analysis formulation would be:

$\hat{x} = \text{argmin}_x \, ||Ax - y||^2 + \lambda ||\mu_\phi(x)||_2^2$
$+\mu||x - \mu_\theta(\mu_\phi(x))||_2^2$

- Any $\hat{x} \in \mathbb{R}^{M \times N}$ potentially reachable
- is working
- Similar to related works [1,2,3]

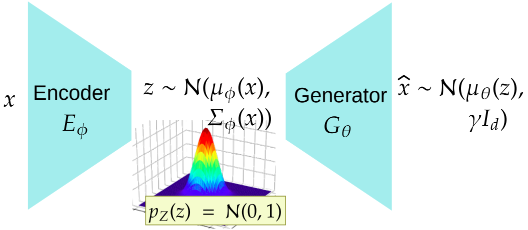


Figure: A variational autoencoder

[1] M. Duff N.D. Campbell, M.J. Ehrhardt, *Regularising Inverse Problems with Generative Machine Learning Models*, arXiv preprint, 2021.
[2] M. González, A. Almansa, P. Tan, *Solving inverse problems by joint posterior maximization with autoencoding prior*, SIAM, 2022.
[3] S.A. Bigdeli, M. Zwicker, P. Tan, *Image restoration using autoencoding priors*, VISIGRAPP, 2018.

[*Contributions*] Test methodology

Image restoration process

Neural network training on ideal images (x_i) $\longrightarrow$ Gradient descent on $||AG(z) - y||^2 + \lambda||z||^2$ for a measured image y

[*Contributions*] Test methodology

Image restoration process

Neural network training on ideal images (x_i) $\longrightarrow$ Gradient descent on $||AG(z) - y||^2 + \lambda||z||^2$ for a measured image y

Neural network settings

- Simple convolutionnal structure, latent dimension denoted as k
- Adam optimiser, $lr = 10^{-4}$, 100 epochs

[*Contributions*] Test methodology

Image restoration process

Neural network training on ideal images (x_i) $\longrightarrow$ Gradient descent on $||AG(z) - y||^2 + \lambda||z||^2$ for a measured image y

Neural network settings

- Simple convolutionnal structure, latent dimension denoted as k
- Adam optimiser, $lr = 10^{-4}$, 100 epochs

Image restoration setting

- 2 inverse problems: denoising (given σ) and inpainting (given p, $\sigma = 10$). σ and p vary depending on the dataset
- 2 metrics: MSE (Mean Squared Error) and SSIM (Structural SIMilarity)

[*Contributions*] Comparison of analysis and synthesis approaches

		Inpainting				Denoising			
		$p = 80\%$		$p = 50\%$		$\sigma = 65$		$\sigma = 25$	
		MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM
MNIST	Synthesis	**0.016**	**0.72**	0.014	**0.75**	**0.014**	**0.75**	0.013	**0.76**
	Analysis	0.017	0.55	**0.010**	0.69	0.014	0.58	**0.007**	0.68
CelebA	Synthesis	**0.0062**	**0.68**	0.0059	0.68	**0.0061**	**0.68**	0.0059	**0.69**
	Analysis	0.0078	0.64	**0.0048**	**0.71**	0.008	0.63	**0.0044**	0.67

Figure: Results of synthesis and analysis approaches on MNIST and CelebA using a VAE. On 4 inverse problems, using 50 test images per problem.

[*Contributions*] Comparison of analysis and synthesis approaches

		Inpainting				Denoising			
		$p = 80\%$		$p = 50\%$		$\sigma = 65$		$\sigma = 25$	
		MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM
MNIST	Synthesis	**0.016**	**0.72**	0.014	**0.75**	**0.014**	**0.75**	0.013	**0.76**
	Analysis	0.017	0.55	**0.010**	0.69	0.014	0.58	**0.007**	0.68
CelebA	Synthesis	**0.0062**	**0.68**	0.0059	0.68	**0.0061**	**0.68**	0.0059	**0.69**
	Analysis	0.0078	0.64	**0.0048**	**0.71**	0.008	0.63	**0.0044**	0.67

Figure: Results of synthesis and analysis approaches on MNIST and CelebA using a VAE. On 4 inverse problems, using 50 test images per problem.

[*Contributions*] Comparison of analysis and synthesis approaches

		Inpainting				Denoising			
		$p = 80\%$		$p = 50\%$		$\sigma = 65$		$\sigma = 25$	
		MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM
MNIST	Synthesis	**0.016**	**0.72**	0.014	**0.75**	**0.014**	**0.75**	0.013	**0.76**
	Analysis	0.017	0.55	**0.010**	0.69	0.014	0.58	**0.007**	0.68
CelebA	Synthesis	**0.0062**	**0.68**	0.0059	0.68	**0.0061**	**0.68**	0.0059	**0.69**
	Analysis	0.0078	0.64	**0.0048**	**0.71**	0.008	0.63	**0.0044**	0.67

Figure: Results of synthesis and analysis approaches on MNIST and CelebA using a VAE. On 4 inverse problems, using 50 test images per problem.

[*Contributions*] Comparison of analysis and synthesis approaches (2)

Figure: Visual results of synthesis and analysis approaches on CelebA for a denoising inverse problem ($\sigma = 25/255$).

- More accurate results on easy inverse problems
- But more artefacts

[*Contributions*] Getting better results with the same network size

The problem is the quality of the generated images

- Increase the dimension of the latent space ? Hard with VAE.

$$\rightarrow \textit{Posterior collapse:} \quad \forall x \in X, q_\phi(z_i|x) = p(z_i)$$

[1] B. Dai, D. Wipf, *Diagnosing and enhancing VAE models*, ICLR 2020

[2] P.Ghoshn, M. Sajjadi, A.Vergari, M.Black, *From Variational to Deterministic Autoencoders*, ICLR 2019

[*Contributions*] Getting better results with the same network size

The problem is the quality of the generated images

- Increase the dimension of the latent space ? Hard with VAE.

$$\rightarrow \textit{Posterior collapse:} \quad \forall x \in X, q_\phi(z_i|x) = p(z_i)$$

$\Rightarrow$ Switch to *regularized* autoencoders [2] (RAE)

Two possibilities:

- Stochastic autoencoder with variable KL-divergence strength parameter[1]
- Deterministic autoencoder with constraints on z[2] ($\rightarrow \lambda||z||_2^2$)

[1] B. Dai, D. Wipf, *Diagnosing and enhancing VAE models*, ICLR 2020

[2] P.Ghoshn, M. Sajjadi, A.Vergari, M.Black, *From Variational to Deterministic Autoencoders*, ICLR 2019

[*Contributions*] Getting better results with the same network size

The problem is the quality of the generated images

- Increase the dimension of the latent space ? Hard with VAE.

$$\rightarrow \textit{Posterior collapse:} \quad \forall x \in X, q_\phi(z_i|x) = p(z_i)$$

$\Rightarrow$ Switch to *regularized* autoencoders [2] (RAE) $\rightarrow$ non generative autoencoders

Two possibilities:

- Stochastic autoencoder with variable KL-divergence strength parameter[1]
- Deterministic autoencoder with constraints on z[2] ($\rightarrow \lambda||z||_2^2$)

[1] B. Dai, D. Wipf, *Diagnosing and enhancing VAE models*, ICLR 2020
[2] P.Ghoshn, M. Sajjadi, A.Vergari, M.Black, *From Variational to Deterministic Autoencoders*, ICLR 2019

[*Contributions*] Comparison VAE / RAE (1)

	Inpainting				Denoising			
	$p = 80\%$		$p = 50\%$		$\sigma = 65$		$\sigma = 25$	
	MSE	SSIM	MSE	SSIM	MSE	SSIM	MSE	SSIM
VAE	0.0062	0.68	0.0059	0.68	0.0061	0.68	0.0059	0.69
RAE	**0.0030**	**0.83**	**0.0012**	**0.90**	**0.0034**	**0.75**	**0.0012**	**0.87**

Table: Metrics for image restoration problems with VAE and RAE using the synthesis approach on CelebA dataset. Latent dimension $k = 64$ for VAE, $k = 200$ for RAE.

- RAE has significant better results than VAE

[Contributions] Comparison VAE / RAE (2)

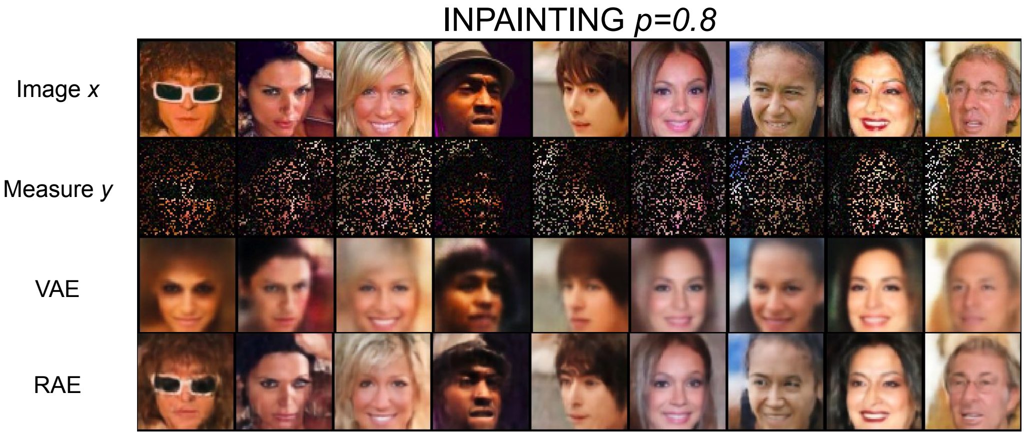


Figure: Visual results using RAE and VAE networks using the synthesis approach, for an inpainting problem ($p = 0.8$ missing pixels, $\sigma = 10/255$)

[*Contributions*] Comparison VAE / RAE (3)

INPAINTING *p=0.8*

Image *x* | Measure *y* | VAE | RAE-l2

Figure: Visual results (zoom)

[*Contributions*] Preliminary results

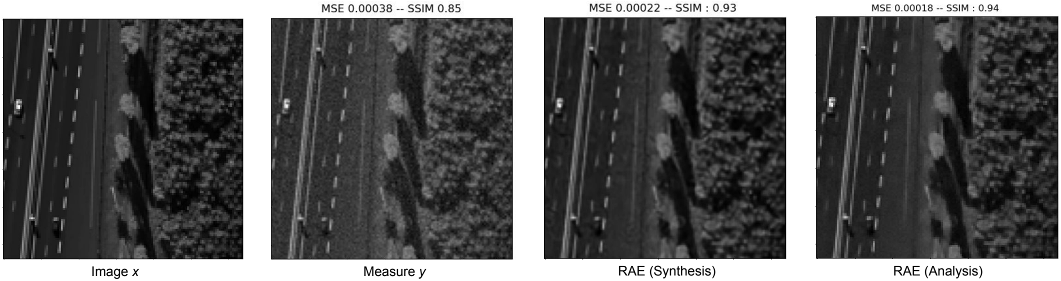


Figure: Results of RAE using analysis and synthesis approaches. On a denoising problem ($\sigma = 5/255$). Images furnished by CNES.

[*Perspectives*] Adaptive regularization

$\text{argmin}_z \, ||AD_\theta(z) - y||_2^2 + \lambda ||z||_2^2 \rightarrow$ not informative enough

- Not image dependent
- Not landscape dependent

[*Perspectives*] Adaptive regularization

$\text{argmin}_z \, ||AD_\theta(z) - y||_2^2 + \lambda||z||_2^2 \rightarrow$ not informative enough

- Not image dependent
- Not landscape dependent

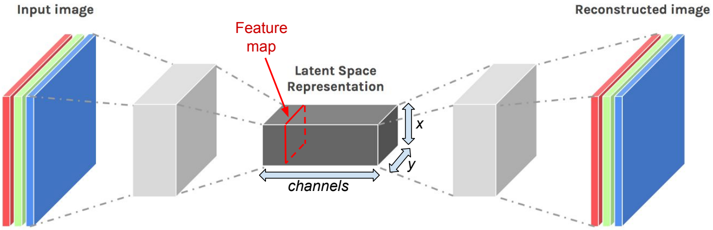


Figure: A fully convolutional autoencoder for translation-invariant representations

[*Perspectives*] Adaptive regularization

$\text{argmin}_z \, ||AD_\theta(z) - y||_2^2 + \lambda||z||_2^2 \rightarrow$ not informative enough

- Not image dependent
- Not landscape dependent

Figure: A restored image

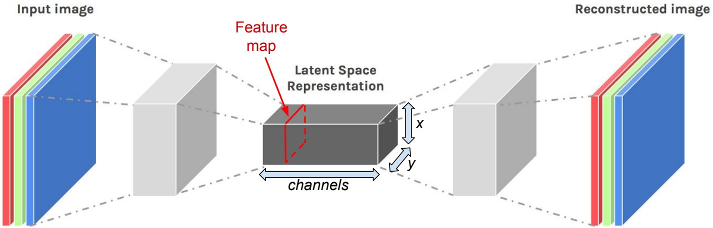


Figure: A fully convolutional autoencoder for translation-invariant representations

[*Perspectives*] Adaptive regularization

$\text{argmin}_z \, ||AD_\theta(z) - y||_2^2 + \lambda||z||_2^2 \rightarrow$ not informative enough

- Not image dependent
- Not landscape dependent

Figure: A restored image

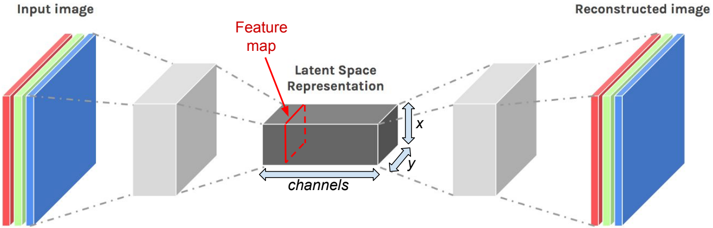


Figure: A fully convolutional autoencoder for translation-invariant representations

[*Perspectives*] Adaptive regularization

$\text{argmin}_z \, ||AD_\theta(z) - y||_2^2 + \lambda||z||_2^2 \rightarrow$ not informative enough

- Not image dependent
- Not landscape dependent

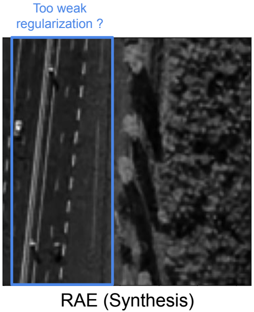


Figure: A restored image

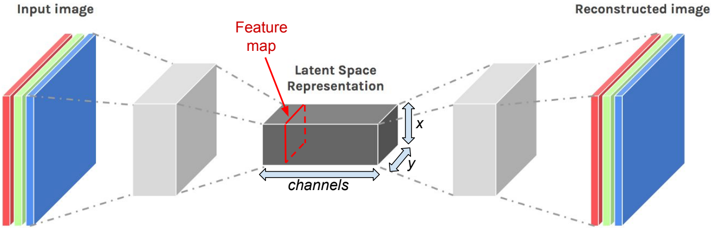


Figure: A fully convolutional autoencoder for translation-invariant representations

 Conclusion

- *Ongoing* work to regularize satellite image restoration problems using (generative) autoencoders
- First results interesting on CelebA, needs further ameliorations for satellite images

[*Conclusion*] Conclusion

- *Ongoing* work to regularize satellite image restoration problems using (generative) autoencoders
- First results interesting on CelebA, needs further ameliorations for satellite images

Thanks for your attention !