Régularisation par modèles pseudo-génératifs pour la restauration d'images

Maud BIQUARD[1,3] Marie CHABERT[2] Florence GENIN[3] Christophe LATRY[3] Thomas OBERLIN[1]

[1]ISAE-Supaero, 10 avenue Edouard Belin, 31400 Toulouse, France

[2]IRIT/INP-ENSEEIHT, 31071 Toulouse, France

[3]CNES, 18 avenue Edouard Belin, 31400 Toulouse, France

Résumé – Cet article s'intéresse aux méthodes de restauration d'images formulées comme des problèmes inverses, dans lesquelles le choix de la régularisation a un impact considérable sur les performances. Les méthodes de régularisation basées sur un réseau de neurones génératif, récemment introduites dans la littérature, se montrent très efficaces mais s'avèrent fortement dépendantes de la qualité du réseau utilisé. Cet article propose d'utiliser un autoencodeur variationnel moins contraint, et d'inférer dans un second temps la distribution des données dans l'espace latent. Les simulations sur les bases de données MNIST et CelebA, pour différents problèmes inverses de difficulté variée, permettent de conclure à un gain en terme de qualité d'image.

Abstract – This paper considers image restoration seen as an inverse problem, for which the regularization has a considerable impact on the performance. Regularization methods based on a generative neural network, recently introduced in the literature, are interesting but prove to be highly dependent on the generative network. This paper proposes to use a generative autoencoder that is less constrained in the latent space. A second step infers the latent distribution learned by the encoder after the training, which is used afterwards as a regularization. The simulations on the MNIST and CelebA datasets for various inverse problems show a gain in terms of image quality.

1 Introduction

La résolution de problèmes inverses est à la base de nombreux algorithmes de restauration d'images. Il s'agit de restaurer une image x à partir de la mesure y et de la connaissance du modèle direct :

$$y = Ax + n \tag{1}$$

où A est l'opérateur de dégradation et n le bruit. Ce problème étant en général mal posé, il est nécessaire de le régulariser au moyen d'une fonction ou pénalité, notée ici R, afin de favoriser les solutions possédant certaines propriétés. On recherche classiquement la solution du problème (1) sous la forme :

$$\hat{x} = \arg\min_x ||y - Ax||^2 + \lambda R(x) \tag{2}$$

La variation totale [17] et les régularisations reposant sur la norme ℓ^p de x sont des régularisations classiques en traitement d'images. Les méthodes dites *plug-and-play* [18] utilisent des débruiteurs comme régularisation implicite du problème inverse.

L'utilisation de réseaux de neurones a permis d'améliorer considérablement les performances des algorithmes de restauration. Une partie de ces méthodes consiste à apprendre à inverser une dégradation spécifique de manière supervisée à partir de données sous forme de couples (image originale, image dégradée) [4]. Ces approches sont généralement très efficaces mais ne sont pas génériques car le modèle appris est spécifique à la dégradation considérée. Au contraire, les méthodes *plug-and-play* sont génériques car l'apprentissage porte uniquement sur le terme de régularisation, ce qui les rend adaptables à n'importe quel problème inverse. En utilisant un réseau de neurones débruiteur, elles atteignent l'état de l'art en restauration d'images [16, 15]. Une autre classe de méthodes pour résoudre (2) utilise les réseaux génératifs [1], qui fournissent un a priori au sens bayésien. Mais les performances deviennent limitées dès lors que le réseau n'est plus parfaitement adapté aux données, ce qui est la motivation principale de ce travail.

2 Régularisation par réseau génératif

Le but des modèles génératifs est de synthétiser des données réalistes $x \in X$, issues d'une distribution p_X inconnue, de telle sorte que :

$$z \sim p_Z \Rightarrow x = G(z) \sim p_X \tag{3}$$

où G est le générateur, z un vecteur issu d'un espace Z appelé espace latent, et p_Z une loi *a priori* sur Z. Il existe plusieurs sortes de modèles génératifs, notamment les autoencodeurs variationnels (VAE, [11]), les réseaux antagonistes[8], les flots normalisants [3] ou plus récemment les modèles de diffusion [9]. Le VAE est un autoencodeur dans lequel le décodeur est utilisé comme modèle génératif. L'apprentissage du VAE cherche à maximiser la vraisemblance marginale, en l'approchant par une borne inférieure dite ELBO (*Evidence Lower BOund*) [11], ce qui revient, en adoptant les hypothèses et les notations de la Figure 1, à maximiser

$$L(\theta, \phi) = -\big[\frac{1}{2\gamma^2}||x - \mu_\theta(z)||^2 + \log\gamma\big] - \mathrm{KL}(q_\phi(z|x)||p_Z(z)) \tag{4}$$

où KL est la divergence de Kullblack-Leibler, $q_\phi(z|x)$ correspond à la distribution variationnelle paramétrée par l'encodeur et $\mu_\theta(z)$ à la sortie du décodeur. Le premier terme de (4) est un

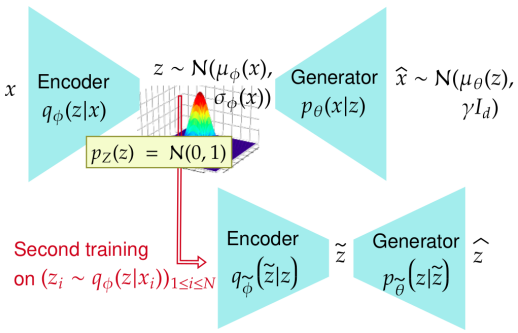

FIGURE 1 : Modèle utilisé pour la régularisation. Le premier autoencodeur est un γ-VAE. Le deuxième est un VAE, appris dans un deuxième temps, sur la base d'entraînement encodée $\mathcal{D}_z$.

terme de fidélité entre l'entrée et la sortie du VAE, alors que le second contraint la distribution variationnelle $q_\phi(z|x)$ à être proche de p_Z. Une fois appris, ces modèles génératifs peuvent être utilisés pour régulariser un problème inverse quelconque, comme proposé par [1] :

$$\hat{z} = \arg\min_z ||AG(z) - y||_2^2 + \lambda ||z||^2 \text{ avec } \hat{x} = G(\hat{z}) \quad (5)$$

où (5) est minimisée par descente de gradient. L'idée est de rechercher la solution $\hat{x}$ la plus proche de la mesure y dans l'espace image du générateur en imposant $x = G(z)$. La régularisation additionnelle $\lambda||z||^2$ vient de la loi *a priori* dans l'espace latent, supposée normale, permettant d'interpréter $\hat{z}$ comme un estimateur du maximum a posteriori dans l'espace latent. Cette méthode se montre particulièrement performante lorsque les données sont aisément représentables sémantiquement (par exemple des visages) et le problème inverse fortement mal posé. Dans le cas contraire, les performances sont moins bonnes car la contrainte $x = G(z)$ restreint trop fortement l'espace de recherche de la solution [1]. Notamment, sur des données présentant une très grande variété, comme des images satellites, cette méthode semble difficilement applicable. D'une part, prétendre apprendre un modèle génératif pertinent sur ces données semble peu raisonnable en terme de quantité de données, de coût calculatoire et donc de consommation énergétique. D'autre part, même avec un tel modèle, la minimisation de (5) pose de nouveaux problèmes du fait de la complexité accrue du générateur. De ce fait, [7, 10] proposent des formulations bayésiennes différentes afin de mieux exploiter l'*a priori* génératif.

Par ailleurs, certains travaux récents [5, 14] relaxent la contrainte $x = G(z)$ en recherchant la solution dans l'espace image X. De cette manière, tout $x \in X$ est théoriquement accessible. Dans la suite, par analogie avec les approches classiques de restauration d'images, on parlera d'approche en synthèse quand la recherche de la solution se fait dans l'espace latent, et d'approche en analyse lorsqu'elle est faite dans l'espace des images.

3 Modèle pseudo-génératif

3.1 γ-VAE

Notre objectif est d'améliorer les performances de [1] quand le modèle génératif atteint ses limites. Au vu de la formulation (5), une question se pose : un modèle génératif est-il vraiment nécessaire ? Dans (5), nous avons constaté que, pour de nombreux problèmes inverses, $\lambda = 0$ fonctionne très bien car la contrainte $x = G(z)$ est souvent suffisante. Par ailleurs, la formulation (5) peut s'utiliser avec un simple autoencodeur, en utilisant le décodeur à la place du modèle génératif. Dans ce cas, la contrainte $x = G(z)$ (où G est le décodeur du réseau) est moins contraignante car l'espace image du décodeur est plus grand. Il devient ainsi vraiment important de contraindre la recherche dans l'espace latent avec $\lambda > 0$.

En pratique, nous avons constaté que l'utilisation d'un autoencodeur non génératif, même bien régularisé [6], ne structure pas l'espace latent aussi efficacement qu'un VAE. Par conséquent, nous introduisons une modification du VAE classique en augmentant la taille de son espace latent afin de générer une plus grande diversité d'images. Cependant, le VAE fait face à un problème particulier dénommé le "*posterior collapse*" [13] : certaines dimensions de l'espace latent sont inactives et ne sont pas utilisées par le décodeur. Ainsi, qu'importe la dimension de l'espace latent, le VAE utilisera uniquement le nombre d'unités latentes nécessaire pour l'équilibre entre le coût de reconstruction et la KL divergence [2].

Pour les VAEs classiques, le paramètre γ, défini Figure 1 et intervenant dans (4), est fixé à 1. Une solution est de considérer $\gamma < 1$, ce qui diminue l'importance de la KL divergence dans la fonction de coût. Plus de dimensions latentes seront alors utilisées, l'autoencodeur sera plus fidèle mais la distribution encodée des données ressemblera moins à l'*a priori* $p_Z = N(0, I_d)$. En pratique, afin d'éviter un réglage manuel de γ, on le considère comme un paramètre du réseau [2]. De cette manière, γ s'adapte automatiquement à la taille de l'espace latent. On appelle ce VAE modifié le γ-VAE.

3.2 Régularisation apprise en deux temps

Le γ-VAE n'est plus un bon modèle génératif, car la valeur faible de γ ne contraint pas suffisamment le modèle à respecter la loi *a priori* dans l'espace latent p_Z. Par conséquent, dans un objectif de restauration d'images (5), le terme $\lambda||z||^2 \propto -\lambda \log p_Z(z)$ n'est plus une régularisation adaptée, nous proposons ici une meilleure alternative.

Si $\mathcal{D} = (x_i)_i$ est la base de données d'entraînement, on note $\mathcal{D}_z = (z_i)_i$ les représentations latentes correspondantes, avec $z_i \sim q_\phi(z|x_i)$. $\mathcal{D}_z$ possède une distribution empirique $\tilde{p}_Z$ qui diffère de $p_Z = N(0, I_d)$ de manière non négligeable. L'idée est maintenant d'estimer $\tilde{p}_Z$ et de remplacer $\lambda||z||^2$ par $-\lambda \log \tilde{p}_Z(z)$ dans (5). Il serait possible d'estimer $\tilde{p}_Z$ à l'aide d'une distribution paramétrique telle qu'une gaussienne, une laplacienne, ou encore un mélange de gaussiennes. Expérimentalement, de tels modèles ne donnent pas de résultats satisfaisants en restauration. En revanche, les réseaux de neurones sont des bons estimateurs de distributions. Nous proposons donc une approche en deux temps. Dans un premier temps, l'entraînement du γ-VAE, qui assure la régularisation du problème inverse, est réalisé avec un *a priori* $p_Z = N(0, I_d)$. Dans un deuxième temps, un VAE auxiliaire de paramètres $(\tilde{\theta}, \tilde{\phi})$ est entraîné sur $\mathcal{D}_z$. Le processus est illustré Figure 1. La fonction de coût $\tilde{L}(z)$ du VAE auxiliaire sera utilisée à la place de $||z||^2$ dans (5) comme terme de régularisation du γ-VAE lors de la restauration d'images. On espère ainsi maximiser la vraisemblance marginale de z en maximisant sa borne inférieure $-\tilde{L}(z)$.

3.3 Formulations en analyse et synthèse de la restauration

Pour l'algorithme de restauration d'images, nous avons testé une formulation en synthèse et une formulation en analyse.

Formulation en synthèse

La formulation en synthèse considérée est une généralisation de l'équation (5) :

$$\hat{z} = \arg\min_z ||A(G(z)) - y||_2^2 + \lambda R(z) \text{ avec } \hat{x} = G(\hat{z}). \quad (6)$$

R est la fonction de régularisation sur l'espace latent. Si on considère que $p_Z = N(0, I_d)$, $R(z) = ||z||^2$ comme dans (5). Un *a priori* plus spécifique, appris dans un deuxième temps, est $R(z) = \tilde{L}(z)$ où $\tilde{L}$ est la fonction de coût associée au deuxième VAE.

Formulation en analyse

En analyse, on considère la formulation suivante :

$$\hat{x} = \arg\min_x ||\mathcal{A}(x) - y||_2^2 + \lambda R(\mu_\phi(x)) + \mu||x - \mu_\theta(\mu_\phi(x))||_2^2 \quad (7)$$

où $\mu_\phi(x)$ est la sortie de l'encodeur, $\mu_\theta(\mu_\phi(x))$ celle de l'autoencodeur. (7) reprend la proposition de [14] mais avec un terme de régularisation supplémentaire $\mu||x - \mu_\theta(\mu_\phi(x))||_2^2$ qui approxime la distance entre x et les données d'entraînement [12, 5]. Ce terme est nécessaire au bon fonctionnement de l'approche en analyse lorsque le réseau utilisé est un VAE.

4 Expérimentations

Dans cette section, nous évaluons les performances du γ-VAE en restauration d'images selon les formulations en synthèse et en analyse et en utilisant pour régularisation la norme 2 et une régularisation $R(z)$ apprise par un deuxième autoencodeur. Ces performances sont comparées à celles de [1].

4.1 Méthodologie

Les expérimentations sont réalisées sur les bases de données MNIST et CelebA. MNIST est composée d'images de chiffres en noir et blanc, de taille 28×28. CelebA est composée d'images RGB de visages de célébrités, de taille 64×64.

Un γ-VAE a été entraîné sur chaque base de données. Un γ-VAE composé uniquement de couches de neurones totalement connectés a été entrainé sur MNIST. La motivation étant à terme de traiter des images naturelles qui nécessitent l'invariance par translation, nous avons également considéré unγ-VAE totalement convolutionnel, c'est-à-dire composé uniquement de couches de convolution. Nous l'avons entrainé sur la base de donnée CelebA, car ce sont des données relativement simples, même si elles ne nécessitent pas d'invariance par translation.

L'espace latent est de taille 200 pour MNIST, et de taille $32\times8\times8$ pour CelebA. Il est important de noter que les espaces latents sont, à dessein, de grande taille comparée aux tailles usuelles que l'on peut trouver pour ces bases de données : les réseaux ainsi obtenus sont de moins bons réseaux génératifs, mais leur espace image est plus grand. Le VAE entraîné dans un deuxième temps sur l'espace latent pour la régularisation a la même structure que le γ-VAE (totalement connecté sur MNIST, totalement convolutionnel sur CelebA), mais avec un nombre de canaux réduit. Le VAE proposé dans [1] est également entraîné sur MNIST et CelebA. Ce VAE a la même structure que le γ-VAE.

Pour le VAE, on utilise la régularisation L2 $R(z) = ||z||^2$ pour la restauration. Pour le γ-VAE, $R(z) = ||z||^2$ et $R(z) = \tilde{L}(z)$ sont testées.

Quatre problèmes de restauration d'images sont considérés : deux problèmes de débruitage avec du bruit blanc gaussien, d'écart-type $\sigma = 25/255$ et $\sigma = 65/255$; deux problèmes d'*inpainting* avec $p = 80\%$ et $p = 50\%$ de pixels manquants aléatoirement répartis sur l'image. Pour ces problèmes, un bruit blanc gaussien d'écart-type $\sigma = 10/255$ est ajouté.

L'évaluation des résultats utilise deux métriques : la MSE (*Mean Squared Error*) et le SSIM (*Structural SIMilarity*). Les réglages des différents paramètres de restauration (λ pour l'approche en synthèse, λ et μ pour l'approche en analyse) sont effectués à l'aide d'une recherche sur grille sur un ensemble de validation.

4.2 Résultats

Le Tableau 1 présente les résultats de restauration d'images en analyse et en synthèse pour le γ-VAE. Tout d'abord, avec $R(z) = ||z||^2$, le γ-VAE est dans la grande majorité des cas significativement meilleur que le VAE, que ce soit en synthèse ou en analyse. Ensuite, l'approche en synthèse donne en moyenne de meilleurs résultats que l'approche en analyse, sauf sur MNIST pour le problème d'*inpainting* avec 50% de pixels manquants. Cela semble pertinent car c'est le plus simple des quatre problèmes inverses considérés. Cependant, il est un peu décevant que l'analyse ne surpasse pas la synthèse sur d'autres problèmes inverses. Elle reste néanmoins compétitive, notamment sur CelebA. Enfin, la régularisation dans un deuxième temps à l'aide d'un VAE semble intéressante. En effet, c'est elle qui donne les meilleurs résultats sur MNIST et sur CelebA, et pour tous les problèmes inverses considérés.

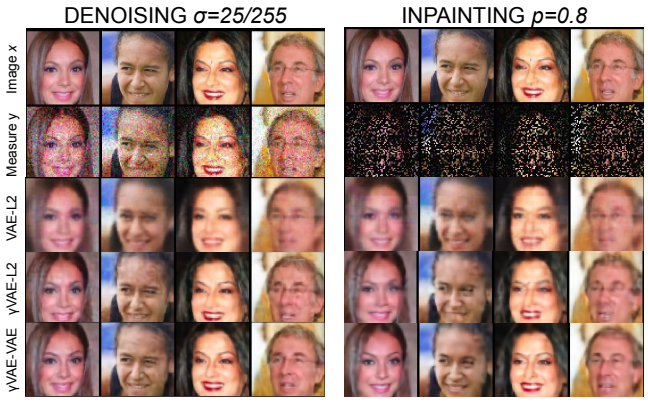

FIGURE 2 : Visualisation des différentes méthodes testées. Restauration en synthèse. VAE ou γVAE est le type de réseau de neurones utilisé. -L2 ou -VAE est le type de régularisation utilisé.

La Figure 2 confirme les résultats du Tableau 1. La restauration avec un VAE est floue : le réseau génératif n'est pas assez performant et donc l'espace image du décodeur est trop restreint. Avec un γ-VAE et une régularisation L2, la restauration est plus nette mais fait apparaître des artefacts. Ceux-ci semblent indiquer que $\tilde{p}_Z$ diffère trop de p_Z et que la régularisation $R(z) = ||z||^2$ n'est pas adaptée. Notre interprétation

Datasets	Inverse Problem	VAE-L2 (Syn)		γVAE-L2 (Syn)		γVAE-VAE (Syn)		VAE-L2 (An)		γVAE-L2 (An)		γVAE-VAE (An)	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MNIST	den25	26.88	0.9638	29.40	0.9809	**29.74**	**0.9839**	22.09	0.6784	27.09	0.9093	27.00	0.9136
	den65	22.98	0.9204	24.49	0.9369	**24.52**	**0.9417**	18.87	0.5753	23.99	0.8942	24.04	0.9036
	inp50	22.37	0.9038	22.73	0.8785	23.26	0.9098	20.69	0.6892	24.86	0.9011	**25.37**	**0.9144**
	inp80	17.47	0.7370	18.88	0.7396	19.98	**0.8282**	16.78	0.6053	19.11	0.7566	**20.07**	0.7865
CelebA	den25	24.93	0.7570	29.98	0.8989	**30.21**	**0.9059**	23.33	0.6805	29.90	0.8960	30.07	0.9004
	den65	23.97	0.7233	25.29	0.7779	**25.82**	**0.8077**	20.90	0.5781	25.34	0.7784	25.79	0.8029
	inp50	24.81	0.7555	30.19	0.9171	**30.49**	**0.9214**	23.31	0.7029	29.97	0.9104	29.42	0.9013
	inp80	24.02	0.7365	25.87	0.8355	**26.63**	**0.8549**	21.37	0.6207	25.72	0.8266	26.43	0.8457

TABLE 1 : Tableau récapitulatif des résultats. den correspond à du débruitage, inp de l'inpainting. VAE-L2 est le VAE avec une régularisation L2, γVAE-L2 le γ-VAE avec une régularisation L2 et γVAE-VAE le γ-VAE avec une régularisation utilisant un 2ème VAE. (Syn) correspond à l'approche en synthèse, (An) à l'analyse. Les meilleures performances pour chaque problème inverse sont en gras.

semble être correcte car la régularisation avec un deuxième VAE pour estimer $\tilde{p}_Z$ permet de supprimer ces artefacts en contraignant davantage la recherche dans l'espace latent. Les résultats obtenus sont alors nets et sans artefacts. La Figure 3

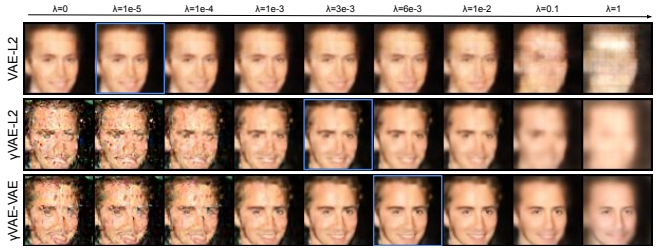

FIGURE 3 : Restauration en fonction de λ pour les différentes méthodes testées. Sur un problème de débruitage ($\sigma = 25/255$) en synthèse. L'image encadrée correspond au niveau de régularisation choisi pour les expérimentations de la Table 1.

donne un aperçu de l'impact de la régularisation R. Pour le VAE, la restauration avec $\lambda = 0$ fonctionne. En effet, comme le VAE est génératif, la contrainte $x = G(z)$ dans la formulation (5) suffit pour régulariser le problème. Cela n'est plus le cas pour le γ-VAE. L'influence de $\lambda R(z)$ est importante et on constate que la régularisation à l'aide d'un deuxième VAE est significativement meilleure.

5 Conclusion

Dans cet article nous avons proposé une méthode de résolution des problèmes inverses rencontrés en restauration d'image, inspirée de [1] mais utilisant un VAE possédant un grand espace latent. Ceci permet d'augmenter la taille de l'espace image du décodeur et donc de l'espace des solutions du problème inverse, en contraignant moins l'espace latent du réseau. Les résultats obtenus ainsi sont meilleurs en terme de métrique, mais comportent des artefacts visuels. L'introduction d'une nouvelle régularisation dans l'espace latent, apprise dans un deuxième temps, permet de supprimer ces artefacts et d'obtenir des résultats intéressants.

Par la suite, il serait pertinent de faire fonctionner cette approche sur des bases de données constituées d'images naturelles pour lesquelles l'architecture totalement convolutionnelle du réseau de neurones est importante. On recherchera également de meilleurs moyens d'estimer la distribution des données dans l'espace latent.

Références

[1] Ashish BORA, Ajil JALAL, Eric PRICE et Alexandros G. DIMAKIS : Compressed sensing using generative models. *International Conference on Machine Learning*, 2017.

[2] Bin DAI et David WIPF : Diagnosing and enhancing vae models. *International Conference of Learning Representations*, 2019.

[3] Laurent DINH, David KRUEGER et Yoshua BENGIO : Nice : Nonlinear independent components estimation. *Proceedings of International Conference of Learning Representations*, 2015.

[4] Chao DONG, Chen Change LOY, Kaiming HE et Xiaoou TANG : Learning a deep convolutional network for image super-resolution. *European Conference on Computer Vision*, pages 184–199, 2014.

[5] Margaret DUFF, Neill D. F. CAMPBELL et Matthias J. EHRHARDT : Regularising inverse problems with generative machine learning models. *arXiv :2107.11191*, 2022.

[6] Partha GHOSH, Mehdi S. M. SAJJADI, Antonio VERGARI et Michael BLACK : From variational to deterministic autoencoders. *Proceedings of International Conference of Learning Representations*, 2020.

[7] Mario GONZÀLEZ, Andrès ALMANSA et Pauline TAN : Solving inverse problems by joint posterior maximization with autoencoding prior. *SIAM Journal on Imaging Sciences*, 15(2):822–859, 2022.

[8] Ian J. GOODFELLOW, Jean POUGET-ABADIE, Mehdi MIRZA, Bing XU, David WARDE-FARLEY, Sherjil OZAIR, Aaron COURVILLE et Yoshua BENGIO : Generative adversarial networks. *Communications of the ACM*, 2014.

[9] Jonathan HO, Ajay JAIN et Pieter ABBEEL : Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 2020.

[10] Matthew HOLDEN, Marcelo PEREYRA et Konstantinos C. ZYGALAKIS : Bayesian imaging with data-driven priors encoded by neural networks. *SIAM Journal on Imaging Sciences*, 15(2):892–924, 2022.

[11] Diederik P KINGMA et Max WELLING : Auto-encoding variational bayes. *International Conference on Learning Representations*, 2014.

[12] Housen LI, Johannes SCHWAB, Stephan ANTHOLZER et Markus HALTMEIER : Nett : solving inverse problems with deep neural networks. *Inverse Problems*, 36:065005, 2020.

[13] James LUCAS, George TUCKER, Roger GROSSE et Mohammad NOROUZI : Understanding posterior collapse in generative latent variable models. *ICLR 2019 Workshop*, 2019.

[14] Thomas OBERLIN et Mathieu VERM : Regularization via deep generative models : an analysis point of view. *In International Conference on Image Processing (ICIP)*, pages 404–408. IEEE, 2021.

[15] Yaniv ROMANO, Michael ELAD et Peyman MILANFAR : The little engine that could : Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10:1804–1844, 2017.

[16] Ernest K. RYU, Jialin LIU, Sicheng WANG, Xiaohan CHEN, Zhangyang WANG et Wotao YIN : Plug-and-play methods provably converge with properly trained denoisers. *International Conference on Machine Learning*, 2019.

[17] A. TIKHONOV : *On the stability of inverse problems*. 1943.

[18] Singanallur V. VENKATAKRISHNAN, Charles A. BOUMAN et Brendt WOHLBERG : Plug-and-play priors for model based reconstruction. *IEEE Global Conference on Signal and Information Processing*, 2013.