Approximation de lois *a posteriori* de problèmes inverses à l'aide d'autoencodeurs variationnels

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

[1]ISAE-Supaero, 10 avenue Edouard Belin, 31055 Toulouse, France

[2]IRIT/Toulouse-INP, 2 rue Charles Camichel, 31000 Toulouse, France

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

Résumé – Cet article s'intéresse à la résolution de problèmes inverses pour la restauration d'images. Il se concentre sur les méthodes qui recherchent leur solution dans l'espace latent d'un réseau de neurones génératif de type autoencodeur variationnel (VAE). Nous nous intéressons en particulier à l'approximation de la distribution *a posteriori* du problème inverse par inférence variationnelle. Cet article présente un cadre général permettant d'estimer cette loi de manière jointe dans l'espace latent et dans l'espace image. Il propose une modélisation de l'incertitude dans l'espace image qui prend en compte la corrélation entre les pixels. Les simulations sur la base d'images CelebA comparent différentes modélisations et démontrent la pertinence de la méthode proposée.

Abstract – This paper focuses on solving inverse problems in image restoration and on methods that seek their solution in the latent space of a generative neural network, specifically a variational autoencoder (VAE). More precisely, we tackle the approximation of the inverse problem's posterior distribution using variational inference. This paper introduces a general framework for estimating this posterior jointly in the image and latent spaces and proposes an uncertainty modeling approach in the image space that accounts for pixel-wise correlations. Simulations on the CelebA dataset compare different modeling strategies and demonstrate the relevance of the proposed approach.

1 Introduction

La restauration des images est une étape préalable à leur exploitation, indispensable dans de nombreux domaines d'application tels que la biologie, l'astrophysique ou l'observation de la Terre. Il s'agit de restaurer une image x à partir de sa mesure bruitée y et de la connaissance du modèle direct :

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

où A est l'opérateur de dégradation et n le bruit. En général, ce problème est mal posé et comporte plusieurs solutions. Une manière de le résoudre consiste à calculer l'estimateur du Maximum *A Posteriori* (MAP), défini par :

$$\max_x p_{X|Y}(x|y) \Leftrightarrow \min_x -\log p_{Y|X}(y|x) - \log p_X(x) \tag{2}$$

où $p_{Y|X}(y|x)$ est la vraisemblance, $p_{X|Y}(x|y)$ la distribution *a posteriori*, et $p_X(x)$ la loi *a priori* de x, traduisant la connaissance sur la solution recherchée à ce problème mal posé.

Les méthodes classiques considèrent un *a priori* ou une régularisation simple, basée par exemple sur la norme l^1 comme pour la variation totale. Cependant, l'utilisation des réseaux de neurones a permis d'améliorer considérablement les performances par rapport aux méthodes classiques. Nous considérons ici les travaux qui apprennent uniquement la régularisation (ou, dans le cadre bayésien, la loi *a priori*), l'avantage principal étant que le modèle peut ensuite être utilisé pour résoudre n'importe quel problème inverse, rendant la méthode très générique. Remarquons que d'autres parties de la littérature utilisent l'apprentissage pour directement inverser le problème, de manière supervisée ou non, avec des architectures classiques ou par algorithmes dépliés (*unfolding*). On peut distinguer deux grandes familles de méthodes pour apprendre la régularisation : la première, dite *plug-and-play*, le fait de manière implicite par des débruiteurs [11]. La seconde famille, que nous nommerons optimisation latente, résout le problème d'optimisation de l'Equation (2) dans l'espace latent d'un modèle génératif. Dans deux travaux récents [1, 2], nous avons proposé une nouvelle méthode d'optimisation latente utilisant des réseaux compresseurs, qui permet en outre d'approcher la distribution *a posteriori* $p_{X|Y}(x|y)$, ce qui est d'une grande importance dans des applications critiques ou scientifiques, où l'estimation des incertitudes associées à une estimation est cruciale. Dans ces travaux, l'approximation de la loi *a posteriori* se fait explicitement, permettant d'échantillonner cette loi de manière très rapide. Dans cet article, nous présentons le cadre général unifiant ces travaux [1, 2], que nous étendons en considérant un décodeur permettant de modéliser les corrélations entre pixels.

2 Pré-requis

Les modèles génératifs à variable latente permettent de synthétiser des données réalistes $x \in X$, issues d'une distribution p_X inconnue, de telle sorte que :

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

où G est le générateur, z un vecteur issu d'un espace Z appelé espace latent, et p une loi *a priori* sur Z. Depuis 2014, l'apprentissage profond a révolutionné les modèles génératifs d'image, avec par exemple les autoencodeurs variationnels [10]), les réseaux antagonistes[8] ou plus récemment les modèles de diffusion [9]. Grâce à ces modèles, on peut

apprendre facilement une loi *a priori* ou une régularisation adaptée à un certain type de données, permettant de résoudre n'importe quel problème inverse.

2.1 Approche MAP-z [3]

Considérons un réseau génératif G entraîné sur une base d'images idéales (*i.e.* n'ayant pas subi de dégradation). Etant donnée une image dégradée y, la méthode, introduite par [3], et que l'on nomme MAP-z, consiste à calculer l'estimateur MAP du problème inverse dans l'espace latent du générateur :

$$\arg\max_z p_{Z|Y}(z|y) \Leftrightarrow \arg\min_z -\log p_{Y|Z}(y|z) - \log p(z), \tag{4}$$

avec $p_{Y|Z}(y|z) = p_{Y|X}(y|G(z))$. Dans le cas d'une loi *a priori* gaussienne et d'un bruit n blanc gaussien de variance σ^2 dans l'Eq.2, cela revient à résoudre $\arg\min_z \frac{1}{2\sigma^2}||AG(z) - y||_2^2 + \frac{1}{2}||z||^2$. La solution du problème inverse est alors $x^* = G(z^*)$ où z^* est la solution de l'Eq.4.

Cette approche est séduisante, car elle permet d'exploiter facilement les récentes avancées des modèles génératifs d'image. Cependant, la solution x^* étant dans l'espace image du générateur, les solutions obtenues peuvent s'avérer décevantes lorsque le modèle génératif est imparfait, ou bien pour des échantillons hors distribution. Des travaux de la littérature ont proposé de s'éloigner légèrement de l'espace image [4] en cherchant le MAP de la loi *a posteriori* jointe sur (x, z) [7].

2.2 Autoencodeur variationnel

Dans cet article, nous nous intéressons à un modèle génératif particulier, l'autoencodeur variationnel (VAE). Le modèle génératif associé est défini par :

$$p_\theta(x, z) = p_\theta(x|z)p(z). \tag{5}$$

où $p(z) = \mathcal{N}(0, I)$ et $p_\theta(x|z) = \mathcal{N}(D_\theta(z), \Sigma_\theta(z))$ est la distribution apprise par le décodeur. La loi *a posteriori* associée est :

$$p_\theta(z|x) = \frac{p_\theta(x|z)p(z)}{\int_z p_\theta(x|z)p(z)dz}. \tag{6}$$

Elle n'est pas calculable en temps raisonnable du fait de l'intégrale. Toutefois, elle peut être approchée par la distribution paramétrique $q_\phi(z|x) = \mathcal{N}(E_\phi(x), \text{diag}(\sigma_\phi(x)))$ apprise par l'encodeur du VAE. Les poids θ et ϕ sont appris à l'entraînement du VAE, en maximisant l'*Evidence Lower BOund* (ELBO), une borne inférieure de la log-vraisemblance $\log p_\theta(x)$ qui s'exprime :

$$\mathcal{L}_{\theta,\phi}(x) = \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x|z)\big] - KL\big(q_\phi(z|x)||p(z)\big). \tag{7}$$

Le premier terme contraint la fidélité de la reconstruction des données par l'autoencodeur. Le second contraint la distribution des données dans l'espace latent à ressembler à $p(z)$.

Les différents modèles de décodeurs. La matrice $\Sigma_\theta(z)$ étant de très grande dimension, l'apprendre sans hypothèse particulière conduit à des instabilités et n'est pas réaliste en termes de coût de calcul. Dans la majorité des cas, on suppose donc $\Sigma_\theta(z) = \gamma^2 I$. Mais c'est une modélisation très

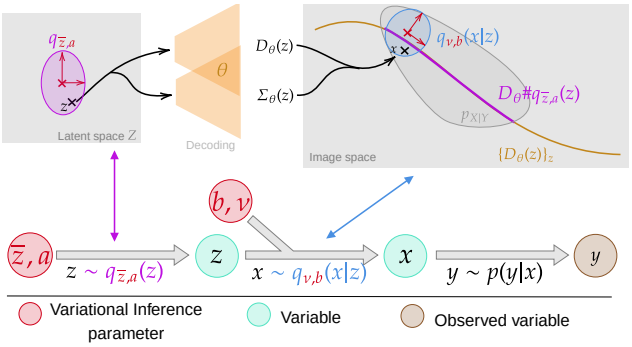

FIGURE 1 : Haut : Illustration de l'échantillonage de la distribution variationnelle. $z \sim q_{\bar{z},a}(z)$ dans l'espace latent, puis $x \sim q_{\nu,b}(x|z)$ dans l'espace image. $D_\theta \# q_{\bar{z},a}$ est la mesure image de $q_{\bar{z},a}$ par D_θ. Bas : Graphe du modèle variationnel proposé.

rudimentaire, où l'erreur de décodage est non corrélée, ce qui pousse généralement à considérer le décodeur comme déterministe : $s = D_\theta(z)$. Une manière de pallier ce problème a été proposée récemment dans [5, 6]. Cette approche entraîne un second décodeur, qui apprend la matrice $L_\theta(z)$ intervenant dans la décomposition de Cholesky $\Sigma_\theta(z)^{-1} = L_\theta(z)L_\theta(z)^T$. La matrice $L_\theta(z)$ est paramétrée de manière creuse, avec l'hypothèse que le coefficient (i, j) est non nul si et seulement si les pixels i et j sont proches dans l'image. Ceci conduit à une matrice diagonale par bande. Cette paramétrisation permet d'apprendre une distribution $p_\theta(x|z)$ réaliste avec un temps de calcul acceptable. En effet, le calcul de la vraisemblance est direct et rapide et l'échantillonnage a une complexité linéaire par rapport au nombre de pixels de l'image. Nous proposons d'exploiter ce modèle pour échantillonner la distribution *a posteriori* d'un problème inverse, comme décrit dans la section suivante.

3 Estimation bayésienne variationnelle

Nous proposons ici d'approcher la loi jointe *a posteriori* $p_{X,Z|Y}(x, z|y)$ par une distribution variationnelle q_ψ issue d'une famille de distributions $\mathcal{E}_\psi$ en résolvant

$$\arg\min_\psi KL(q_\psi(x, z)||p_{X,Z}(x, z|y)). \tag{8}$$

Nous proposons de choisir des distributions cohérentes avec celles utilisées dans le VAE, c'est-à-dire gaussiennes, avec des moyennes qui représentent le code latent $\bar{z}$ ou l'image $D(\bar{z}) + \nu$ les plus probables, et des paramètres a et b qui modélisent l'incertitude, dans les espaces latent et image :

$$\begin{aligned}\mathcal{E}_\psi = \{&q_\psi(x, z) = q_{\nu,b}(x|z)q_{\bar{z},a}(z)|\bar{z}, a \in \mathbb{R}^{C\times M\times N},\\ &b, \nu \in \mathbb{R}^{C'\times M'\times N'}, a, b > 0\}\end{aligned} \tag{9}$$

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

$$q_{\nu,b}(x|z) = \mathcal{N}(x; D_\theta(z) + \nu, \Sigma_{b,\theta}(z))$$

avec $\Sigma_{b,\theta}(z) = L_\theta(z)^{-T}\text{diag}((b_{i,j,k}^2))L_\theta(z)^{-1}$ et L_θ introduit dans Section 2.2. Une illustration de q_ψ est proposée en Figure 1. Résoudre l'Equation 8 revient, de la même manière qu'à l'entraînement d'un VAE, à maximiser l'ELBO :

$$\begin{aligned}&\arg\max_{\bar{z},a,\nu,b} \mathbb{E}_{q_b(x|z)q_{\bar{z},a}(z)}\big[\log p_{Y|X}(y|x)\big]\\ &- KL(q_{b,\nu}(x|z)||p_\theta(x|z)) - KL(q_{\bar{z},a}(z)||p(z))\end{aligned} \tag{10}$$

Les KL-divergences sont exprimables explicitement car les distributions considérées sont gaussiennes. Après simplification, l'Eq. 10 revient à résoudre :

$$\min_{\bar{z},a,\nu,b} \mathbb{E}_{q_b(x|z)q_{\bar{z},a}(z)}\big[\log p_{Y|X}(y|x) + \sum_k (\frac{1}{2}(b_k^2 - 2\log b_k) + \nu^T \Sigma_{b,\theta}(z)\nu)\big] + \sum_j \frac{1}{2}(a_j^2 + \bar{z}_j^2 - 2\log a_j - 1). \quad (11)$$

L'Equation 11 est minimisée par descente de gradient en approximant le gradient de l'espérance par son estimateur *Stochastic Gradient Variational Bayes* (SGVB) [10]. En pratique, comme lors de l'entraînement d'un VAE, un seul échantillon est utilisé à chaque itération pour le calcul de l'estimateur SGVB.

Cas particuliers. Les distributions introduites à l'Eq. 9 sont gaussiennes de moyennes et variances définies afin de ressembler au modèle génératif $p_\theta(x, z)$. Si on souhaite modéliser la loi *a posteriori* uniquement dans l'espace image ou l'espace latent, il est également possible de supposer que $q_{\bar{z},a}(z)$ ou $q_{\nu,b}(x|z)$ sont des Diracs. Notamment, lorsque $q_{\bar{z},a}(z) = \delta(\bar{z}), q_{\nu,b}(x|z) = \delta(D_\theta(z))$, on retrouve la formulation MAP-z de l'Eq. (4). Lorsque $q_{\nu,b}(x|z)$ est un Dirac, l'incertitude est modélisée uniquement dans l'espace latent. C'est ce qui est présenté dans [1] avec l'algorithme Variational Bayes Latent Estimation (VBLE). Lorsque $\Sigma_\theta(z) = \text{diag}(\sigma_\theta(z)_k^2)$, alors le modèle de loi *a posteriori* dans l'espace image ne prend pas en compte les corrélations entre pixels. C'est l'algorithme VBLE-xz introduit dans [2].

4 Expérimentations

Dans cette section, nous évaluons la méthode présentée dans la Section 3 avec différentes distributions variationnelles.

4.1 Méthodologie

Les expérimentations sont réalisées sur la base de données CelebA, composée d'images de célébrités de taille 64×64, en noir et blanc. Un VAE a été entraîné sur cette base de données. Ce VAE est composé uniquement de couches totalement convolutionnelles, la motivation à terme étant de traiter des images naturelles. D'abord, l'encodeur E_ϕ et le décodeur D_θ sont entraînés en considérant le modèle génératif $p_\theta(x|z) = \mathcal{N}(D_\theta(z), \gamma^2 I)$ avec γ un paramètre du réseau. Dans un deuxième temps, le décodeur de variance Σ_θ est appris avec la même paramétrisation que [5], les autres paramètres du réseau étant fixés, en minimisant $\mathbb{E}[-\log p_\theta(x|z)]$.

On évalue les performances de différentes méthodes. VBLE-xz est le cas général introduit dans la Section 3, tandis que VBLE-xz ($\nu = 0$) est le cas où $q_{\nu,b}(x|z) = \mathcal{N}(x; D_\theta(z), \Sigma_{b,\theta}(z))$. Ensuite, VBLE-z [1] consiste à approcher la loi *a posteriori* du problème inverse uniquement dans l'espace latent, en considérant $q_{\nu,b}(x|z) = \delta(D_\theta(z))$. VBLE-x consiste à approcher la loi *a posteriori* uniquement dans l'espace image, en considérant $q_{\bar{z},a}(z) = \delta(\bar{z})$ et $\nu = 0$. Enfin, MAP-z est la méthode de référence, consistant à minimiser l'Eq. 4.

Les différentes méthodes sont évaluées sur trois problèmes inverses : défloutage , super-résolution d'un facteur 2, et *inpainting* avec 50% de pixels manquants aléatoirement répartis sur l'image. Pour le défloutage, du bruit blanc gaussien d'écart-type $\sigma = 2.55/255$ ou $7.65/255$ est rajouté. Pour la super-résolution et l'inpainting, on considère $\sigma = 1/255$.

Les métriques utilisées pour évaluer les performances sont le rapport signal-sur-bruit-crête (PSNR), et la *Learned Image Patch Similarity* (LPIPS) [12]. De plus, pour évaluer la qualité de l'approximation de la loi *a posteriori*, des *Interval Coverage Probabilities* (ICPs) de niveau α sont utilisés. Cela représente la proportion de la vraie erreur, par pixel, contenue dans le α-quantile de l'erreur prédite. Ainsi, plus l'ICP est proche de α, plus les incertitudes sont pertinentes.

4.2 Résultats

TABLE 1 : Tableau des résultats. Deblur (Gaussian) est du défloutage avec un noyau gaussien de variance 1, et du bruit blanc gaussien de d'écart-type $\sigma = 7.65$. SISR est de la super-résolution. ICP correspond à l'ICP de niveau $\alpha = 0.9$.

	Deblur (Gaussian)			**SISR** $\times 2$			**Inpainting (random)**		
	PSNR	LPIPS	ICP	PSNR	LPIPS	ICP	PSNR	LPIPS	ICP
MAP-z	33.25	0.090	x	33.28	0.0755	x	33.90	0.057	x
VBLE-z	33.29	0.089	0.74	**33.36**	0.0750	0.71	34.03	0.056	0.62
VBLE-x	33.21	0.9478	0.64	33.24	0.0751	0.075	33.34	0.062	0.64
VBLE-xz ($\nu = 0$)	33.13	0.090	**0.87**	33.18	0.077	0.86	33.50	0.060	0.73
VBLE-xz	**34.04**	**0.072**	0.86	33.30	**0.070**	**0.88**	**34.92**	**0.042**	**0.78**

La Table 1 montre des résultats quantitatifs des différentes méthodes, introduites dans la Section 4.1 précédente, sur les problèmes inverses testés. En termes de PSNR et LPIPS, les résultats de VBLE-x et VBLE-xz ($\nu = 0$) sont un peu moins bons que ceux des autres méthodes. MAP-z et VBLE-z ont des métriques assez proches, et VBLE-xz est supérieur à toutes les autres méthodes : cela confirme qu'autoriser une petite déviation dans l'espace image permet de s'éloigner de l'espace image du générateur et ainsi d'être plus précis sur la solution proposée. Ensuite, en termes d'ICPs, celles de VBLE-z et VBLE-x sont assez faibles. Par conséquent, il semble que modéliser l'incertitude uniquement dans l'espace latent ou l'espace image n'est pas suffisant pour capturer toute la complexité de la loi *a posteriori* du problème inverse. Les ICPs des deux versions de VBLE-xz sont bien plus proches de la valeur théorique 0.9, cela prouve la pertinence de considérer un modèle joint dans les espaces latent et image pour approximer la loi *a posteriori*.

La Figure 2 montre des résultats visuels pour un problème de défloutage et d'*inpainting*. La limite de VBLE-z est visible sur le problème d'inpainting : la modélisation de l'incertitude uniquement dans l'espace latent ne permet pas de modéliser les incertitudes très structurées propres à l'inpainting. En revanche, on retrouve le masque d'*inpainting* dans les écart-types marginaux des autres méthodes, ce qui montre la pertinence de la modélisation utilisée, et notamment du facteur b permettant d'adapter la variance du décodeur à un problème donné. De plus, les échantillons générés par les méthodes modélisant une partie de l'incertitude sur x sont significativement plus réalistes et plus nets. Cela montre l'intérêt de la distribution variationnelle considérée, en particulier, le fait de considérer une distribution $q_{\nu,b}(x|z)$ avec de la corrélation entre pixels. Enfin, la méthode VBLE-xz est celle qui offre les meilleures

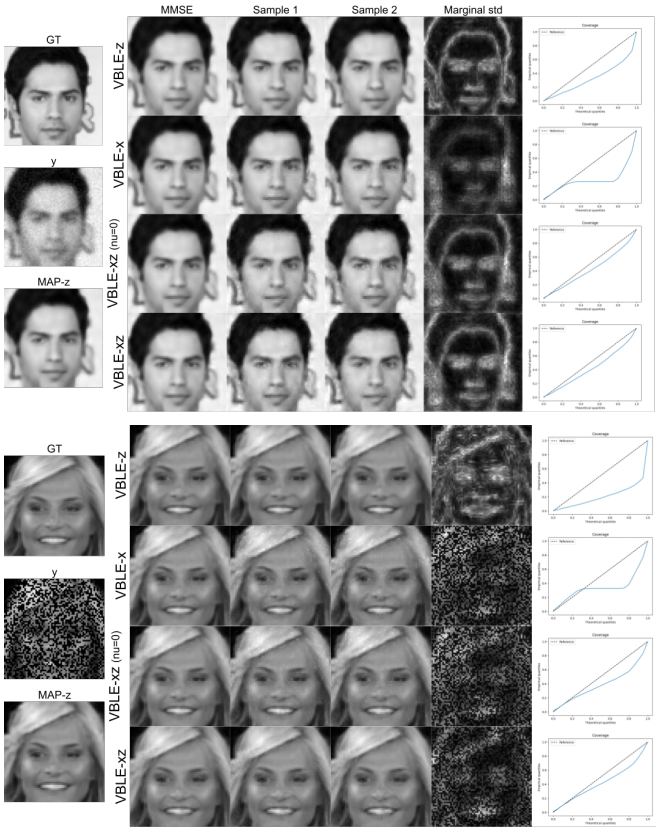

FIGURE 2 : Visualisation des différentes méthodes testées en défloutage et *inpainting*. Le MMSE et les écart-types marginaux sont obtenus en prenant la moyenne et l'écart-type de 100 échantillons de la distribution *a posteriori*. A droite, des courbes de *coverage* représentent l'ICP(α) en fonction de α.

performances visuelles : l'estimateur MMSE est plus net, les échantillons sont réalistes, et les incertitudes cohérentes.

Ainsi, ces simulations montrent l'intérêt de VBLE-xz, qui possède de meilleurs résultats que toutes les autres variantes testées, autant en termes de PSNR/LPIPS que de qualité des incertitudes et de la visualisation.

5 Conclusion

Dans cet article, nous avons proposé une méthode permettant d'approcher la loi *a posteriori* d'un problème inverse de manière conjointe dans les espaces latent et image grâce à un VAE. Cette approche repose sur une modélisation explicite des incertitudes, et permet d'échantillonner efficacement la loi *a posteriori*. Les simulations montrent l'intérêt de modéliser cette loi dans les espaces latent et image de manière jointe, par rapport à la modéliser uniquement dans l'un ou l'autre espace car les incertitudes prédites sont bien meilleures dans le premier cas. De plus, la prise en compte d'une corrélation entre pixels dans l'espace image permet d'obtenir des échantillons très réalistes de la loi *a posteriori*.

Nous avons montré sur des images naturelles [1] et satellites [2] que l'approche VBLE permet une bonne restauration et une bonne estimation des incertitudes, avec un temps de calcul de l'ordre de celui du MAP-z, restant même compétitive par rapport aux modèles de diffusion, bien plus coûteux. L'intérêt d'une matrice de covariance non diagonale restera à évaluer dans ce cadre plus réaliste.

Références

[1] Maud BIQUARD, Marie CHABERT, Florence GENIN, Christophe LATRY et Thomas OBERLIN : Variational bayes image restoration with compressive autoencoders. *arXiv preprint arXiv :2311.17744*, 2023.

[2] Maud BIQUARD, Marie CHABERT, Florence GENIN, Christophe LATRY et Thomas OBERLIN : Deep priors for satellite image restoration with accurate uncertainties. *arXiv preprint arXiv :2412.04130*, 2024.

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

[4] Manik DHAR, Aditya GROVER et Stefano ERMON : Modeling sparse deviations for compressed sensing using generative models. *In International Conference on Machine Learning (ICLR)*, pages 1214–1223. PMLR, 2018.

[5] Garoe DORTA, Sara VICENTE, Lourdes AGAPITO, Neill DF CAMPBELL et Ivor SIMPSON : Structured uncertainty prediction networks. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5477–5485, 2018.

[6] Margaret AG DUFF, Ivor JA SIMPSON, Matthias Joachim EHRHARDT et Neill DF CAMPBELL : VAEs with structured image covariance applied to compressed sensing MRI. *Physics in Medicine & Biology*, 68(16):165008, 2023.

[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 (NeurIPS)*, 2020.

[10] Diederik P KINGMA et Max WELLING : Auto-encoding variational bayes. *In International Conference on Learning Representations (ICLR)*, 2014.

[11] 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 (GlobalSIP)*, 2013.

[12] Richard ZHANG, Phillip ISOLA, Alexei A. EFROS, Eli SHECHTMAN et Oliver WANG : The unreasonable effectiveness of deep features as a perceptual metric. *In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.