

Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of some Toulouse researchers and makes it freely available over the web where possible.

This is an author's version published in: https://oatao.univ-toulouse.fr/28453

Official URL: https://doi.org/10.1109/ICIP42928.2021.9506138

To cite this version :

Oberlin, Thomas and Verm, Mathieu Regularization via Deep Generative Models: an Analysis Point of View. (2021) In: 2021 IEEE International Conference on Image Processing (ICIP), 2021 - 2021 (Anchorage, United States).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

REGULARIZATION VIA DEEP GENERATIVE MODELS: AN ANALYSIS POINT OF VIEW

Thomas Oberlin and Mathieu Verm

ISAE-SUPAERO, Université de Toulouse, 10 Avenue Edouard Belin, Toulouse 31400, France

ABSTRACT

This paper proposes a new way of regularizing an inverse problem in imaging (e.g., deblurring or inpainting) by means of a deep generative neural network. Compared to end-to-end models, such approaches seem particularly interesting since the same network can be used for *many* different problems and experimental conditions, as soon as the generative model is suited to the data. Previous works proposed to use a *synthesis* framework, where the estimation is performed on the latent vector, the solution being obtained afterwards via the decoder. Instead, we propose an *analysis* formulation where we directly optimize the image itself and penalize the latent vector. We illustrate the interest of such a formulation by running experiments of inpainting, deblurring and super-resolution. In many cases our technique achieves a clear improvement of the performance and seems to be more robust, in particular with respect to initialization.

***Index Terms*—** inverse problems, regularization, generative models, deep regularization, data-driven priors

1. INTRODUCTION

Inverse problems are ubiquitous in imaging, because many image acquisition pipelines involve degradation operators that need to be inverted afterwards, such as radial projections in tomography or spatial invariant blur in any image or signal acquisition sensor [1]. This task can be seen as an estimation problem, which is often solved via optimization algorithms or Monte Carlo sampling [2]. To circumvent the ill-posedness of the problem, practicioners need *priors* or *regularizers* that will promote some behavior of the solution.

Famous generic priors have been proposed in imaging in the past decades, such as Total Variation [3] and variants [4], which favour piecewise-smooth images. This is done by minimizing the ℓ_1 norm of some linear differentiation operator of the image. If we know a priori a basis in which the data is sparse, the ℓ_1 norm can also be placed on the representation coefficients in that basis, which remains an efficient technique in some applications [5]. More recently, significant improvements have been achieved by considering *plug-and-play priors* and variants [6, 7], often based on highly efficient denoisers such as BM3D [8].

But in many situations such a basis is not known, although it can be *learned* from some available training data. This idea has been first introduced with the problem of dictionary learning [9], which simultaneously learns the dictionary and the noise-free image. It has then been applied to many inverse problems by training the dictionary beforehands on a representative dataset. More recently, numerous works considered to train a deep neural networks (DNN) instead of a dictionary, which appeared to be a more efficient and effective model for images. Some works proposed to learn a denoising networks [10, 11] and use it as a plug-and-play prior; or to learn a 1-class classification DNN [12] which is used as a projection operator onto the set of natural images.

We will focus on a different approach introduced by [13], which simply learns a generative DNN and use it to constrain the solution to live in its range. In this work, authors showed recovery results similar to the ones obtained in compressed sensing, and demonstrated the effectiveness of the approach with a variational autoencoder (VAE) and a Deep Convolutional Generative Adversarial Network (DCGAN). Impressive results have been obtained by considering another kind of generative model termed Glow [14], which is invertible and thus enables to better control the distribution of the latent factor by optimizing the likelihood. Our aim is to show that with such a network, the result can be dramatically improved by optimizing the image itself instead of the latent code, the network being used only for regularization purpose. We name our approach deep *analysis* regularization, as opposed to the *synthesis* regularization proposed in [13], and following the formulation used in sparse recovery [15, 16].

Note that apart from the "deep plug-and-play approaches" discussed above, which use deep learning only in the regularization, there are several lines of works that tackle inverse problems with deep learning in very different ways. To cite a few, this includes unrolled deep networks [17], unsupervised approaches [18] or other end-to-end approaches [19]. Each line of methods has its pros and cons, and comparing those techniques is beyond the scope of the current paper. In a nutshell, plug-and-play approaches seem more generic (one prior for different problems) and somehow more grounded (we can control the optimization algorithm), but they might be outperformed by problem-dependant methods [20].

Part of this work has been funded by the Institute for Artificial and Natural Intelligence Toulouse (ANITI) under grant agreement ANR-19-PI3A-0004.

The remainder of this paper is structured as follows. Section 2 recalls some background and the works upon which we build our analysis deep regularization technique, presented in Section 3. The experimental setting and the corresponding results are then presented in Sections 4 and 5, respectively, while a summary and some perspectives are discussed in Section 6 which concludes the paper.

2. BACKGROUND

2.1. Problem statement

Many imaging problems consist in sensing the image $\mathbf{y}$ from an underlying true scene $\mathbf{x}^*$ according to

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{n}, \tag{1}$$

where $\mathbf{A}$ is some observation operator which is often known with good accuracy, and $\mathbf{n}$ an error term accounting for random effects and potential model mismatch. This model encompasses denoising (A is the identity matrix), inpainting (A is a mask), deblurring (A is Toeplitz), etc. The noise term is often assumed to be independent and Gaussian, although in many imaging modalities is contains a mixing of Gaussian and Poisson components [21].

The inverse problem consists in estimating the true scene $\mathbf{x}$, which amounts to inverting the linear operator $\mathbf{A}$. This problem is in general ill-posed, because $\mathbf{A}$ is often singular or at least badly conditioned. To overcome this, one needs to bring additional information by means of a *regularizer* φ. The formulation of the inverse problem then writes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\varphi(\mathbf{x}) \tag{2}$$

where parameter λ tunes the level of regularization.

Such a formulation can be related to Bayesian estimation: if the noise $\mathbf{n}$ is assumed to be Gaussian, then (2) is the maximum a posteriori (MAP) estimate under the prior $p(\mathbf{x}) \propto e^{-\varphi(\mathbf{x})}$. Note that this remains valid if the noise follows a different statistics, as long as we replace the least-square by the adequate divergence measure. For convenience, we will restrict our study to the least-square formulation, which is the most convenient and the most widely used.

2.2. Priors via deep generative neural networks

Auto-encoders are special kinds of neural networks. They are composed of an encoder which computes the latent code $\mathbf{z} = E(\mathbf{x})$, and a decoder or a generator $D(\mathbf{z})$, both being trained so that $\mathbf{x}_i \approx D(E(\mathbf{x}_i))$ for all data samples $\mathbf{x}_i$ of a given training dataset. While autoencoders have been introduced a long time ago, there has been a stunning renewal with the introduction of so-called variational auto-encoders (VAEs) in [22, 23]. The aim of these works was to constrain the latent space to exhibit a proper structure, which can be seen as a regularization of the learning problem. The first purpose was to be able to generate realistic images by sampling from a simple distribution in the latent space. Recall that the loss function of a VAE is intractable and requires the use of approximate inference techniques. To overcome this, some works [24, 25] followed a different approach inspired by normalizing flows, where each layer of the network remains invertible, its gradient being efficiently computed to allow for backpropagation. Once learned, such networks can be use as priors to regularize any inverse problem, as described in the following paragraph.

2.3. Synthesis-based regularization

We describe here our baselines [13, 14]. They both assume that a generative model has been learned beforehand on a representative database. Then, they propose to solve (2), by constraining the sought image $\mathbf{x}$ to be in the range of the generator: $\mathbf{x} = D(\mathbf{z})$ for some latent vector $\mathbf{z}$. Both works add an extra regularizer, based on the normal distribution assumed for the latent vector, which finally gives

$$\hat{\mathbf{x}} = D\left(\arg\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{A}D(\mathbf{z}) - \mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_2^2\right). \tag{3}$$

3. AN ANALYSIS FORMULATION

3.1. Limits of the synthesis regularization

The synthesis formulation (3) operates in the latent space, which brings some clear benefits but also several drawbacks. On the one hand, the solution of such a problem is likely to be visually consistent, since it has been generated by the decoder. But on the other hand, the found solution is often very different from the ground truth, because of the strong non-convexity of the problem which makes the solution highly dependent to initialization. To reduce this effect one should properly initialize the optimization algorithm, for instance by setting $\mathbf{z}_0 = E(\mathbf{x}_0)$, with $\mathbf{x}_0$ an initial guess. According to our experience, this trick however does not always produce an image which is fully consistent with the observations.

3.2. An analysis formulation

To circumvent the drawbacks discussed above, we propose here a straightforward solution which consists in optimizing directly the image. The regularization is performed in the latent space, capitalizing on the Gaussian distribution which is assumed for the latent vector. This writes:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|E(\mathbf{x})\|_2^2. \tag{4}$$

We named this formulation *analysis*. Note that, contrary to the synthesis formulation, the role of parameter λ is fundamental, since it controls the only regularization brought about by the network.

The two formulations are the counterpart of what was already studied concerning sparse representations in redundant dictionaries [15]. But in the case of deep regularization, both approaches always lead to different results, even if the network is invertible, because of the strong non-convexity of the problem. The remainder of the paper is devoted to illustrate the pros and cons of the two formulations, on various simple inverse problems. To this end, we describe the experimental setting and the results in the next two sections.

4. EXPERIMENTAL SETTING

We restrict our study to the comparison between the proposed analysis formulation and the more standard synthesis, as implemented by [14]. Note that this recent work competes favorably with several state of the art techniques, we thus tried to reproduce a close experimental setting.

For the generative DNN, we use the Glow network as introduced in [25]. Glow is inspired by flow-based generative models [26], whose particularity is to exhibit a tractable log-likelihood for the generative process, which avoids the use of approximate inference techniques. This is achieved by means of a particular architecture, composed of invertible layers with (fast) tractable Jacobian. Instead of standard convolutions, the layers of Glow are thus composed of split, 1×1 convolutions and actnorm operations, we refer to [25] for more details. We trained Glow on the CelebA dataset [27], composed of colored images of faces with size 64×64. We used 4 blocks with 32 steps of flow each and additive coupling layers. The optimizer chosen for the training was Adam with a learning rate of 10^{-4} ($\beta_1 = 0.9$ & $\beta_2 = 0.999$).

Insofar as Glow is invertible, it can be used both as an Encoder $E(.)$ or a Decoder $D(.)$ for (respectively) the analysis and the synthesis formulations. Because of the Gaussian prior on the latent space, the *regularizer* φ will be the squared ℓ_2 norm. The optimisation problems (3) and (4) are solved by gradient descent (either on the latent space or on the image). The initialization of the gradient descent is done, as explained by [14], by $\mathbf{z}^0 = 0$ and $\mathbf{x}^0 = D(0)$ for the synthesis and analysis formulations, respectively. Parameter λ is chosen empirically around 10^{-4} for each method and every inverse problems, so as to obtain the best performance.

In the experiments, we consider the three following inverse problems:

- Super-resolution: increasing the resolution of an image which has been previously downsampled by a factor 2 or 4 with local averaging (uniform filter).
- Deblurring: removing the blur caused by a 7×7 uniform filter.
- Inpainting: filling the gaps in an image caused by the application of a mask. We consider two different settings, either 60% of uniformly random missing pixels or a squared mask centered on the image of size $\frac{1}{9} \times \frac{1}{9}$.

The metrics used to evaluate the effectiveness of the result are the widely used PSNR and SSIM. For the sake of reproducibility, our code and additional results are available at `https://github.com/Mathieu-Verm/Analysis_Synthesis_GLOW`.

5. RESULTS

We run experiments on 50 images from the test set of CelebA. Let us first show visual comparisons between the synthesis and analysis frameworks, to highlight the pros and cons of both formulations. The result of the deblurring and the super-resolution experiments are depicted on Fig. 1 for three images of the test set, that were randomly selected. Concerning the deblurring, the results of the synthesis method exhibit artefacts, that are particularly visible in the background and do not disappear if we increase λ. The analysis formulation instead shows a far better result, both for the background and the face. To explain this, we believe that the synthesis approach struggles to represent the background, in particular for the first image because its pattern is a bit unusual and might not have been seen during training. In the analysis approach instead there is no need to generate the background, since the optimization takes place in the image space. The super-resolution experiment investigates a more difficult setting where fewer information is available, and for this the synthesis approach might seem preferable. Indeed, the faces reconstructed by synthesis are much more pleasant than the analysis counterpart, which suffer from strong artefacts (the prior does not seem to work well enough, even for a large value of λ). Note however that the synthesis results, although visually better, are outperformed by analysis in terms of PSNR or SSIM, as we will see later.

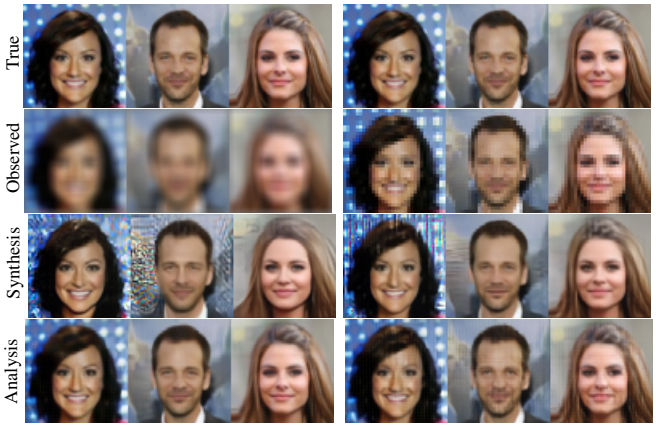


Fig. 1. Results of deblurring (left) and 2x super-resolution (right) for three images of the test set. From top to bottom: truth, degraded image, results of synthesis and analysis formulations.

Let us now move to the inpainting results, depicted on

Task	PSNR (synthesis)	PSNR (analysis)	SSIM (synthesis)	SSIM (analysis)
Deblurring	23.38 $\pm$2.04	**32.16** $\pm$1.56	0.74 $\pm$0.09	**0.94** $\pm$0.01
Super-resolution (x2)	22.26 $\pm$4.21	**31.19** $\pm$1.33	0.76 $\pm$0.12	**0.93** $\pm$0.01
Super-resolution (x4)	18.94 $\pm$2.81	**24.12** $\pm$1.21	0.61 $\pm$0.11	**0.76** $\pm$0.03
Inpainting (random mask)	21.84 $\pm$3.57	**27.89** $\pm$2.24	0.71 $\pm$0.14	**0.87** $\pm$0.05
Inpainting (structured mask)	**30.40** $\pm$2.53	27.50 $\pm$3.26	**0.94** $\pm$0.02	0.91 $\pm$0.03

Table 1. Average performance over a test set of 50 images and corresponding standard deviations. The best score between analysis and synthesis are highlighted in bold, for both metrics.

Figure 2 for the random and the structured masks. In the first case, the analysis technique seems to achieve the best results, while it is outperformed by its synthesis counterpart for the structured mask. We have the same explanations as in the previous experiment: we believe that in general, the analysis formulation stays closer to the true solution because it does not suffer from the intrinsic bias given by the generative network, and its optimization seems safer (there might be more spurious local minima in the latent space than in the image space). However when the problem becomes too difficult, the analysis does not seem to regularize enough, and the synthesis formulation obtains visually better results, although not clearly better in terms of PSNR or SSIM.

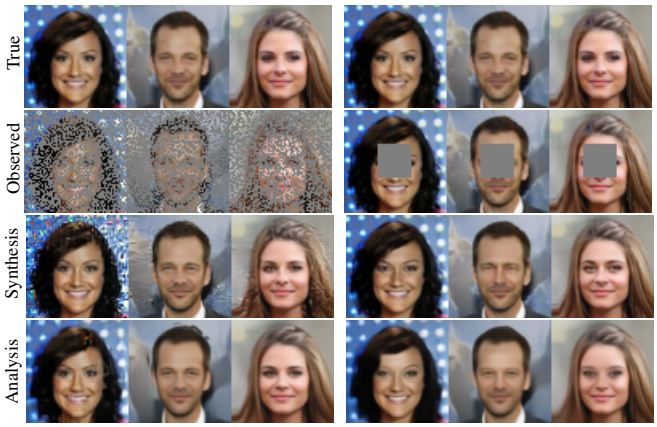


Fig. 2. Inpainting with random mask (left) and structured mask (right), for the same three images of the test set.

We also computed the PSNR and SSIM for the 50 images and the 5 different inverse problems, and gathered the results in Table 1. It confirms the previous findings: in most cases, the analysis formulation enables an image recovery which is dramatically closer to the ground truth than with the synthesis formulation. But for the most difficult problem, i.e., inpainting with a structured mask (and to a lower extent, super-resolution with factor 4), the synthesis formulation seems better. Another interesting point is the variance of the metrics across the 15 images, which is significantly lower for the analysis formulation. This means that the optimization landscape seems somehow safer in the image space that in the latent space, which is an expected benefit of the analysis formulation.

6. CONCLUSION

We presented here a variant of the deep learning regularization via generative models, were the solution is searched directly in the image space. According to our experiments, our *analysis* formulation obtains an estimation of the image which is more robust and often closer to the ground truth, even if sometimes visually less pleasant, particularly when the observation does not contain enough information.

Further work is required to better understand what is the best formulation for a given setting, and to extend these findings to improved learning strategies such as in [28] or [29]. Concerning the applications, the main bottleneck up to now is the limited ability of generative networks to represent heterogeneous datasets, such as images with various size and content. Forthcoming contributions in generative models should soon overcome this, opening a huge number of possible real-world applications.

7. REFERENCES

[1] M. Bertero, *Introduction to inverse problems in imaging*, CRC press, 2020.

[2] J. Idier, *Bayesian approach to inverse problems*, John Wiley & Sons, 2013.

[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[4] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[5] E. Monier, T. Oberlin, N. Brun, X. Li, M. Tencé, and N. Dobigeon, "Fast reconstruction of atomic-scale STEM-EELS images from sparse sampling," *Ultramicroscopy*, p. 112993, 2020.

[6] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction,"

in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2013, pp. 945–948.

[7] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[10] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 3929–3938.

[11] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[12] J.H. Rick Chang, C.-L. Li, B. Poczos, B.V.K. Vijaya Kumar, and A. C. Sankaranarayanan, "One network to solve them all–solving linear inverse problems using deep projection models," in *IEEE International Conference on Computer Vision (CVPR)*, 2017, pp. 5888–5897.

[13] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning (ICML)*, 2017, pp. 537–546.

[14] M. Asim, A. Ahmed, and P. Hand, "Invertible generative models for inverse problems: mitigating representation error and dataset bias," in *International Conference on Machine Learning (ICML)*, 2020.

[15] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse problems*, vol. 23, no. 3, pp. 947, 2007.

[16] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[17] C. Bertocchi, E. Chouzenoux, M-C. Corbineau, J-C. Pesquet, and M. Prato, "Deep unfolding of a proximal interior point method for image restoration," *Inverse Problems*, vol. 36, no. 3.

[18] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9446–9454.

[19] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 5505–5514.

[20] S. Arridge, P. Maass, O. Öktem, and C-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[21] M. Makitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 91–103, 2012.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *International Conference on Learning Representations (ICLR)*, 2014.

[23] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *International Conference on Machine Learning (ICML)*, 2014.

[24] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, "Improved variational inference with inverse autoregressive flow," in *Advances in neural information processing systems (NeurIPS)*, 2016, pp. 4743–4751.

[25] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *Advances in neural information processing systems (NeurIPS)*, 2018, pp. 10215–10224.

[26] L. Dinh, D. Krueger, and Y. Bengio, "Nice: Non-linear independent components estimation," *arXiv preprint arXiv:1410.8516*, 2014.

[27] Z. Liu, P. Luo, X. Wang, and X. Tang, "Large-scale celebfaces attributes (celeba) dataset," *Retrieved August*, vol. 15, pp. 2018, 2018.

[28] M. Gonzalez, A. Almansa, M. Delbracio, P. Musé, and P. Tan, "Solving inverse problems by joint posterior maximization with a VAE prior," *arXiv preprint arXiv:1911.06379*, 2019.

[29] A. Jalal, L. Liu, A. G. Dimakis, and C. Caramanis, "Robust compressed sensing of generative models," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.