

# Apprentissage mono-image contraint pour la super-résolution

Pierrick CHATILLON<sup>1,2</sup>, Yann GOUSSEAU<sup>2</sup>, Sidonie LEFEBVRE<sup>1</sup>

<sup>1</sup>DOTA, ONERA, Université Paris-Saclay, F-91123 Palaiseau, France

<sup>2</sup>Télécom Paris (IDS), 19 place Marguerite Perey, 91120 Palaiseau, France

pierrick.chatillon@onera.fr, yann.gousseau@telecom-paris.fr, sidonie.lefebvre@onera.fr

**Résumé** – Les méthodes basées sur l’apprentissage profond pour la super-résolution (SR) ont été récemment l’objet de nombreux travaux. En particulier, plusieurs articles ont montré que l’étape d’apprentissage peut être effectuée sur une seule image (approches dites internes). La méthode SinGAN est l’une de ces contributions, où la distribution des patches de l’image est apprise sur l’image en entrée et propagée à des échelles plus fines. Il existe des situations dans lesquelles des a priori statistiques peuvent être supposés pour l’image finale. En particulier, de nombreux phénomènes naturels produisent des images ayant un spectre de Fourier décroissant selon une loi de puissance, comme les nuages et autres images de texture. Dans cet article, nous montrons comment de telles informations a priori peuvent être intégrées dans une approche de super-résolution interne, en contraignant la procédure d’échantillonnage de SinGAN. Nous démontrons par diverses expériences que ces contraintes sont effectivement satisfaites, mais aussi que certaines mesures de qualité peuvent être améliorées par l’approche proposée.

**Abstract** – Deep learning based methods for super-resolution (SR) have drawn a lot of attention lately. In particular, various papers have shown that the learning stage can be performed on a single image, resulting in the so-called internal approaches. The SinGAN method is one of these contributions, where the distribution of image patches is learnt on the image at hand and propagated at finer scales. Now, there are situations where some statistical a priori can be assumed for the final image. In particular, many natural phenomena yield images having power law Fourier spectrum, such as cloud and other texture images. In this work, we show how such a priori information can be integrated into an internal super-resolution approach, by constraining the learned up-sampling procedure of SinGAN. We demonstrate on various experiments that these constraints are indeed satisfied, but also that some quality measures can be improved by the proposed approach.

## 1 Introduction

La super-résolution a récemment considérablement progressé grâce à l’utilisation de réseaux de neurones convolutifs, depuis la première architecture SRCNN [3], en passant par de nombreuses méthodes améliorant le PSNR, RDN [14] ou le récent SwinIR [8]. Ces méthodes reposent généralement sur un entraînement sur de grandes bases de données, et sont appelées ”méthodes externes”. En outre, et en profitant de la puissance d’apprentissage des réseaux adversariaux génératifs (GAN), des méthodes ont été proposées pour entraîner une architecture de super-résolution sur une seule image, notamment par l’approche SinGAN [12]. Une telle méthode (appelée ”interne”) peut être considérée comme une extension par les réseaux neuronaux des principes généraux qui sous-tendent les approches par patches de traitement d’image classique.

Dans ce travail, nous nous intéressons à l’application des méthodes de super-résolution internes à un type spécifique d’images, des images de type ”bruit coloré”, c’est-à-dire les images dont le spectre de puissance décroît comme une fonction de puissance  $1/f^\beta$ . Un exemple courant en est le bruit rose, également connu sous le nom de bruit en  $1/f$ . Les processus de bruit coloré sont particulièrement abondants dans la nature [7], et constituent un modèle raisonnable pour les images de nuages, qui constituent notre principal cas d’application. Le but est ici de générer des champs nuageux de grande taille,

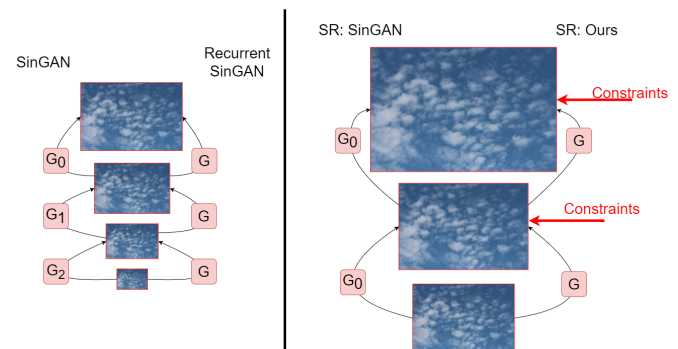


FIGURE 1 – Schéma de l’entraînement (gauche) et de la procédure de SR (droite) pour SinGAN et SinGAN récurrent. L’ajout de contraintes permet de stabiliser la SR.

réalistes et à haute résolution, que ce soit dans le contexte de l’évaluation des performances des capteurs optiques ou pour étalonner les modèles de transfert radiatif. Les simulations d’écoulements turbulents permettent de simuler des champs nuageux réalistes, mais ils reposent sur des modèles stochastiques coûteux basés sur un cadre fractal ou de Fourier. Ils exploitent les propriétés d’invariance d’échelle observées dans les nuages réels : le spectre de puissance du logarithme de leurs propriétés optiques est caractérisé par une pente spectrale d’en-

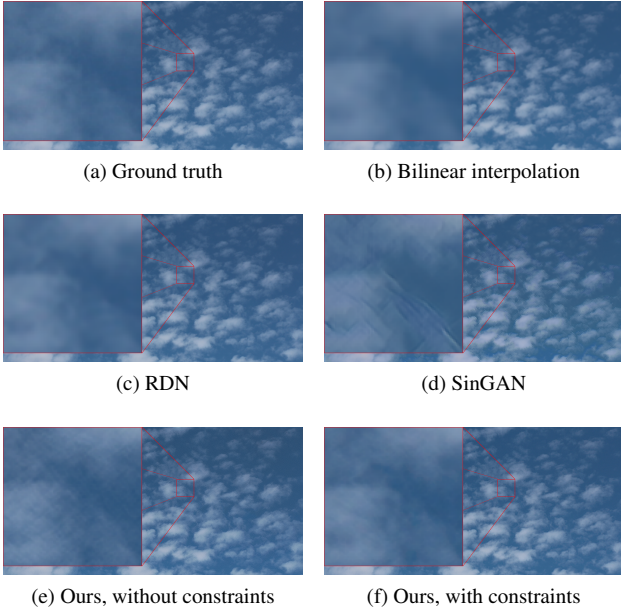


FIGURE 2 – Comparaison visuelle de toutes les méthodes étudiées pour une image tirée de la base Cloud.

viron  $\frac{5}{3}$  [13] de la petite échelle (quelques mètres) à l'échelle extérieure (environ cent kilomètres). Le but de ce travail est de développer une méthode de SR dans l'esprit des approches génératives, permettant d'augmenter la résolution des images de bruit coloré tout en préservant des propriétés statistiques physiquement pertinentes telles que le spectre de puissance. A notre connaissance, un tel modèle n'a pas été proposé dans la littérature, à l'exception de [2], où un modèle multifractal est exploité pour la super-résolution d'images du soleil.

Dans la partie expérimentale de cet article, Section 3, nous montrerons que les contraintes choisies améliorent en outre des mesures perceptuelles et de texture.

## 2 Méthode

### 2.1 Contraintes statistiques

Dans cette section, nous décrivons les contraintes statistiques considérées pour la SR. Nous désignons par  $u$  l'image à basse résolution (LR) et  $w$  le résultat de la super-résolution par un facteur 4. La notation  $\xi$  correspond à la variable de fréquence dans le domaine de Fourier.

**Contraintes spectrales.** Comme expliqué dans l'introduction, nous nous intéressons aux images dont le spectre décroît comme une fonction puissance. Nous imposons que la loi du spectre soit celle d'un champ gaussien stationnaire, sauf aux basses fréquences qui sont conservées. Plus précisément, la projection  $proj_{spectre}$  s'écrit :

$$\begin{cases} proj_{spectre}(\hat{w})_{\xi} & = \hat{w}_{\xi} & \text{if } r \leq r_0 \\ proj_{spectre}(\hat{w})_{\xi_i^r} & = \frac{\hat{w}_{\xi_i^r}}{\|\hat{w}_{\xi_i^r}\|} Q_{\beta_r}^{\mathcal{R}}\left(\frac{i}{|\mathcal{C}_r|}\right) & \text{if } r > r_0 \end{cases} \quad (1)$$

en notant, pour  $r > r_0$ ,  $\xi_i^r$  les fréquences situées sur le cercle  $\mathcal{C}_r$ , ordonnées selon la magnitude de  $\|\hat{w}_{\xi}\|$ , et  $Q_{\beta}^{\mathcal{R}}$  la fonction quantile de la distribution de Rayleigh de paramètre  $\beta$ . Le paramètre de la distribution de Rayleigh,  $\beta_r$ , est choisi de sorte que le module de  $\hat{w}$  ait la moyenne souhaitée sur  $\mathcal{C}_r$  :

$$\beta_r = \sqrt{\frac{2}{\pi}} \frac{\mathbb{E}_{\mathcal{C}_{r_0}}(\|\hat{w}\|)}{\mathbb{E}_{\mathcal{C}_r}(\|\hat{w}\|)} \left(\frac{r}{r_0}\right)^{-p}$$

De plus, pour éviter les artefacts et ne contrôler que la partie pertinente du spectre, nous nous appuyons sur la décomposition "périodique plus smooth" d'une image [10].

**Contrainte d'histogramme.** Afin de contrer certaines limitations du générateur, nous voulons que l'histogramme de couleur de  $w$  corresponde à celui de  $u$ . Nous adoptons une approche de transport optimal par tranche ('sliced') [11]. Cette approche consiste à modifier pas à pas l'histogramme de couleur de  $w$  le long de directions aléatoires  $\theta \in \mathbf{R}^3$  de sorte que les histogrammes projetés sur  $\theta$  de  $u$  et  $w$  coïncident. Nous appelons  $proj_{hist}(w)$  le résultat de ce processus.

**Contrainte de réversibilité.** La condition de réversibilité est un test de cohérence entre l'image d'entrée à basse résolution  $u$  et le résultat de la super-résolution  $w$ . Nous voulons que  $u$  soit le résultat d'un filtrage passe-bas suivi d'un sous-échantillonnage de  $w$ . Cette condition peut être considérée comme un terme de fidélité, rendant le résultat  $w$  plus proche, au sens  $L_2$ , de la vérité terrain. En d'autres termes, nous voulons que notre méthode offre un bon compromis perception-distorsion ([1]). Cette condition peut être calculée par :

$$\mathcal{L}_{rev} = \frac{1}{2} \|D_f(w * g) - u\|^2 \quad (2)$$

où  $D_f$  est l'opérateur de décimation (de la dimension de  $w$  à la dimension de  $u$ ), et le filtre de lissage  $g$  est choisi comme gaussien d'écart type  $\sigma = .7 * f$ , qui est une taille raisonnable au regard du théorème de Shannon. Nous appelons  $proj_{rev}(w)$  l'image résultante d'une courte descente de gradient de  $\mathcal{L}_{rev}$ .

### 2.2 SinGAN

Afin de pouvoir s'adapter à la spécificité statistique d'une seule image, nous avons choisi comme base l'architecture classique de SinGAN.

Rappelons la brièvement : une image d'apprentissage unique est transformée en une pyramide multi-résolution  $\{x_N, \dots, x_0\}$  où  $x_0$  est l'image originale, et  $x_n$  est  $x_0$  sous-échantillonnée  $n$  fois par un facteur  $\frac{4}{3}$ . SinGAN est basé sur une architecture pyramidale de générateurs  $\{G_N, \dots, G_0\}$ . Chaque générateur  $G_i$  est entraîné à agrandir l'image et ajouter stochastiquement du détail de manière à reproduire la distribution des patches de  $x_i$ .

L'itération du réseau  $G_0$  (préconisée par les auteurs de SinGAN) qui n'a jamais été entraîné à recevoir sa propre sortie en entrée, conduit à d'importants artefacts de couleur et de texture. Nous résolvons ce problème en entraînant un générateur récurrent à la manière de [5], puis en appliquant les contraintes décrites en 2.1 à chaque étape de la procédure itérative de SR (comme illustré en Fig.1, droite).

## 2.3 Notre approche

Nous entraînons un unique générateur  $G$  résiduel sur la même pyramide multi-échelle que SinGAN, avec 3 échelles.

Nous alternons l'application de notre générateur récurrent  $G$  et 3 étapes de projection alternée sur nos 3 contraintes comme décrit dans l'algorithme 1 et illustré dans la Fig. 1.

```

Input:  $x_0$ 
 $x = x_0$ 
while zoom désiré non atteint do
     $x = G(x)$ ; // agrandit et ajoute les
    détails
    for step in range(3) do
         $x = proj_{spectre}(x)$ 
         $x = proj_{hist}(x)$ 
         $x = proj_{rev}(x)$ 
    end
end

```

**Algorithm 1:** Algorithme de projection alternée utilisé pour appliquer les contraintes à l'image à chaque échelle.

## 3 Expériences

Nous utilisons 3 bases de données pour nos expériences :

- Une base de données synthétique de 10 images de bruit coloré de  $1200 \times 1200$ . Nous choisissons une pente spectrale de  $\beta = 1, 7$ .
- Une base d'images appelée Cloud de 12 photos de nuages  $1728 \times 1152$  prises depuis le sol.
- Le jeu de données BSD100, d'images naturelles utilisé par les auteurs de SinGAN pour évaluer leur méthode SR.

### 3.1 Métriques d'évaluation

Nous considérons différents indices de qualité : PSNR, NIQE [9], et  $D_{texture}$ . L'indice NIQE utilise les caractéristiques de distribution extraites sur les patches pour évaluer la qualité d'une image. Contrairement au PSNR, il s'agit d'une mesure sans référence. Comme cela a été fait dans l'article [6], et inspiré par l'article [4], nous utilisons une distance de texture  $D_{GATYS}$ . Cette distance varie beaucoup en magnitude selon les images. Afin de calculer des statistiques significatives sur plusieurs images, nous normalisons cette distance pour chaque image par la distance  $D_{GATYS}(I_{GT}, I_{bilinear})$  entre l'image de référence  $I_{GT}$  et le zoom bilinéaire 4x de l'image  $I_{bilinear}$  :

$$D_{texture}(I_{GT}, I) = \frac{D_{GATYS}(I_{GT}, I)}{D_{GATYS}(I_{GT}, I_{bilinear})} \quad (3)$$

Cette distance étalonée peut être considérée comme un ratio de performance par rapport à l'interpolation bilinéaire.

Toutes les expériences sont réalisées avec un zoom 4x. Nous comparons nos résultats à l'interpolation bilinéaire et à RDN [14], une méthode de SR externe visant à maximiser le PSNR. Certains résultats peuvent être comparés visuellement dans la Fig. 2, avec une section  $100 \times 100$  de l'image agrandie par un facteur 8.

## 3.2 Étude d'ablation

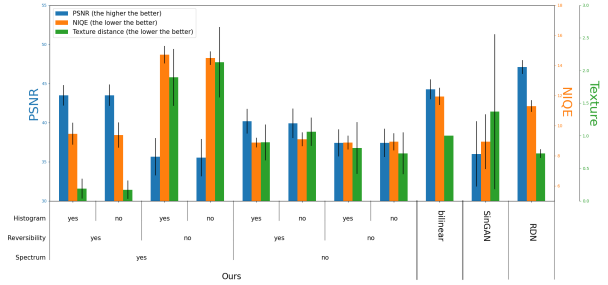


FIGURE 3 – Etude d'ablation réalisée sur l'ensemble des données Cloud. Les performances de l'interpolation bilinéaire, SinGAN et RDN sont représentées à des fins de comparaison.

Nous vérifions tout d'abord que le choix d'un générateur récurrent permet d'améliorer les résultats de SinGAN sur des données auto-similaires, telles que des nuages. Nous remarquons en effet une amélioration du PSNR et de la distance de texture par rapport à SinGAN lorsque nous utilisons notre générateur, **sans appliquer aucune contrainte**.

Nous effectuons ensuite une étude d'ablation sur les différentes contraintes statistiques considérées dans cet article. La Fig. 3 rapporte les performances de la méthode sur le jeu de données Cloud lorsque chaque contrainte est utilisée ('oui') ou non ('non'). Nous observons que, comme prévu, l'utilisation de la contrainte de réversibilité améliore le PSNR, et l'utilisation à la fois de la réversibilité et du spectre donne le meilleur score de texture et le meilleur PSNR de tous les choix.

### 3.3 Analyse quantitative

Nous proposons de comparer l'interpolation bilinéaire, RDN, SinGAN et notre méthode en utilisant les métriques et les ensembles de données décrits ci-dessus.

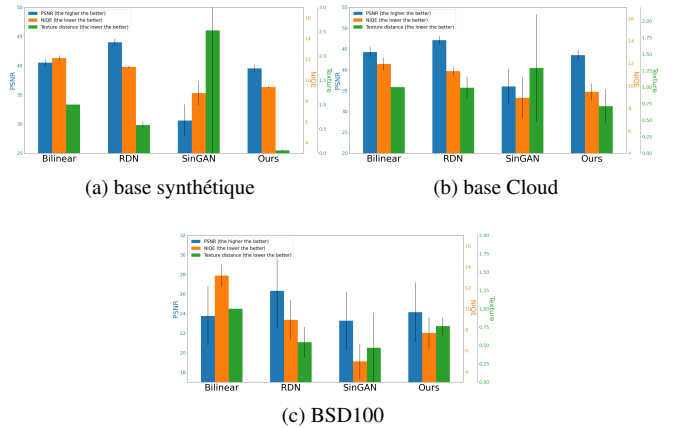


FIGURE 4 – Comparaison des méthodes sur différentes bases pour nos 3 métriques : PSNR, NIQE et distance de texture

**Base synthétique.** Nous vérifions que notre méthode fonctionne bien sur les images de bruit coloré. Nous utilisons des images

de notre ensemble de données synthétiques, nous les sous-échantillons par un facteur 4, puis nous évaluons plusieurs méthodes de zoom de facteur 4. Nous soulignons que notre méthode suppose une pente spectrale de 1.7 pour sa contrainte spectrale, et ces images sont créées pour avoir la même pente de 1.7, donnant ainsi beaucoup d’informations a priori à notre méthode, qui surpasse effectivement les autres en qualité de texture tout en maintenant un bon PSNR (voir Fig.4a). Hormis des scores NIQE comparables aux nôtres, les performances de SinGAN sont particulièrement faibles sur ce jeu de données.

**Cloud et BSD100.** Sur la base BSD100 (Fig. 4c), SinGAN donne de meilleurs résultats pour les deux métriques de qualité perceptuelle, ce qui est logique puisque les images naturelles qui forment le jeu de données BSD100 sont très différentes des images de bruit colorées. Notre a priori spectral n’est pas valide sur cette base et explique cette dégradation. En revanche, sur le jeu de données des nuages (Fig. 4b), nous obtenons de meilleurs résultats en termes de texture et de PSNR que SinGAN, tout en surpassant RDN et l’interpolation bilinéaire sur les deux métriques perceptuelles.

Nous ne prétendons pas rivaliser en termes de PSNR avec une méthode externe telle que RDN (Fig. 4) optimisée pour la performance PSNR. Notre objectif est de respecter les contraintes statistiques, et ce faisant, nous obtenons de bons résultats perceptuels, tout en maintenant des scores PSNR acceptables.

### 3.4 Respect des contraintes

	Bilinéaire	RDN	SinGAN	Ours
Distance d’histogramme couleur	39 ± 28	17 ± 3	938 ± 2202	47 ± 40
Erreur de réversibilité ( $\cdot 10^{-3}$ )	1.2 ± 0.4	0.5 ± 0.2	20.4 ± 49.8	0.6 ± 0.3
Erreur de décroissance spectrale ( $\cdot 10^{-3}$ )	8.0 ± 1.2	8.6 ± 1.3	4.6 ± 1.2	3.3 ± 0.9

TABLE 1 – Évaluation du respect des 3 contraintes pour les 4 méthodes sur la base Cloud

Nous vérifions (Tableau I) sur le jeu de données Cloud que le résultat de notre méthode respecte les contraintes que nous avons fixées avec des métriques étroitement liées à ces contraintes.

## 4 Conclusion

Nous avons introduit une méthode de SR interne qui préserve certaines contraintes statistiques et notamment une propriété de décroissance du spectre de Fourier. L’intérêt de notre approche est démontré sur des résultats de comparaison avec SinGAN, que nous améliorons en termes de PSNR et de fidélité de la texture sur des jeux de données d’images de bruit coloré.

**Remerciements.** Ce travail est soutenu par l’Agence d’Innovation de la Défense et le projet ANR-19-CE40-0005 MISTIC.

## Références

[1] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on com-*

*puter vision and pattern recognition*, pages 6228–6237, 2018.

- [2] P. Chainais, E. Koenig, V. Delouille, and J.-F. Hochedez. Virtual Super Resolution of Scale Invariant Textured Images Using Multifractal Stochastic Processes. *Journal of Mathematical Imaging and Vision*, 39(1) :28–44, 2011.
- [3] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2) :295–307, 2015.
- [4] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015.
- [5] X. He and Z. Fu. Recurrent singan : Towards scale-agnostic single image gans. In *Proceedings of the 2021 5th International Conference on EITCE*, EITCE 2021, page 361–366, New York, NY, USA, 2021. Association for Computing Machinery.
- [6] A. Houdard, A. Leclaire, N. Papadakis, and J. Rabin. Wasserstein Generative Models for Patch-based Texture Synthesis. In *Scale Space and Variational Methods in Computer Vision*, volume LNCS 12679, pages 269–280, 2021.
- [7] M. S. Keshner. 1/f noise. *Proceedings of the IEEE*, 70(3) :212–218, 1982.
- [8] J. Liang et al. Swinir : Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021.
- [9] A. Mittal, R. Soundararajan, and A. Bovik. Making a “completely blind” image quality analyzer. *Signal Processing Letters, IEEE*, 20 :209–212, 03 2013.
- [10] L. Moisan. Periodic plus smooth image decomposition. *Journal of Mathematical Imaging and Vision*, 39(2) :161–179, Feb 2011.
- [11] J. Rabin, J. Delon, and Y. Gousseau. Regularization of transportation maps for color and contrast transfer. In *2010 IEEE International Conference on Image Processing*, pages 1933–1936. IEEE, 2010.
- [12] T. R. Shaham, T. Dekel, and T. Michaeli. Singan : Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4570–4580, 2019.
- [13] F. Szczap et al. A flexible three-dimensional stratocumulus, cumulus and cirrus cloud generator (3dcloud) based on drastically simplified atmospheric equations and the fourier transform framework. *Geosci. Model Dev.*, 7 :1779–1801, 2014.
- [14] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.