

A priori synthétique pour la restauration d'images

Raphaël Achddou, Yann Gousseau, Saïd Ladjal

Abstract—Les réseaux de neurones profonds ont récemment surpassés les méthodes classiques de restauration d'images reposant sur des a priori de régularité. Cependant, ces réseaux nécessitent de grandes bases d'apprentissage, et s'adaptent mal à de nouvelles modalités. Nous proposons d'entraîner ces réseaux sur des bases d'images synthétiques, générées à partir d'un modèle feuilles mortes, nécessitant peu de paramètres et pouvant générer une quantité arbitraire de données. Nous montrons que, pour une tâche de débruitage, les performances de ces réseaux sont très proches de ceux entraînés sur des images naturelles.

English : Deep neural networks have recently surpassed classical image restoration methods based on a priori assumptions. However, these networks require large training bases, and generalize poorly to new modalities. In this paper, we propose to train these networks on randomly generated image bases from a dead leaf model, requiring few parameters and able to generate an arbitrary amount of data. The performances of these networks are very close to those trained on natural images, for Gaussian denoising. After presenting the dead leaf model, we will present our experimental results.

I. INTRODUCTION

Les méthodes classiques de restauration d'images reposent généralement sur des hypothèses de régularité des images. Par exemple, Les méthodes reposant sur la variation totale [1] supposent une distribution laplacienne pour le gradient de l'image, et les méthodes non locales [2] reposent sur une hypothèse d'auto-similarité.

Récemment, les réseaux de neurones profonds (CNN) ont permis d'obtenir des résultats impressionnants pour toutes les tâches de restauration d'images : débruitage, super-résolution mono-image, déconvolution, etc. Ces réseaux ne reposent pas sur des a priori statistiques explicites, nécessitent un entraînement sur des bases de données volumineuses et doivent être ré-entraînés pour chaque nouvelle modalité ou chaque appareil imageur [3]. Dans ce travail, nous montrons que ces réseaux peuvent être entraînés efficacement à partir de bases de données d'images synthétiques basées sur un modèle mathématique dépendant de quelques paramètres. À notre connaissance, il s'agit du premier exemple d'entraînement entièrement synthétique pour les réseaux de restauration d'images naturelles.

Parmi les modèles mathématiques disponibles pour les images naturelles, nous montrons que le modèle feuilles mortes (reposant sur la modélisation de l'occultation et une invariance par changement d'échelle), est suffisant pour atteindre des performances proches de l'état de l'art pour le débruitage gaussien. De plus, nous montrons que ce modèle peut être efficacement combiné avec des bases de données d'images naturelles pour améliorer la capacité des réseaux neuronaux

profonds à préserver les détails, sans altérer leurs performances en terme de PSNR.

Nous pensons qu'une telle étude permet à la fois d'éclairer la manière dont les CNN peuvent résoudre les problèmes de restauration et d'ouvrir des perspectives intéressantes.

Tout d'abord, ce résultat montre que la structure de ces réseaux est adaptée aux tâches de restauration d'images et que, malgré leur grand nombre de paramètres, ils peuvent être rendus quasi-optimaux à partir de quelques principes et hyper-paramètres. Ce résultat, et le fait que des modèles d'images plus simples et moins structurés ne peuvent atteindre des performances de restauration satisfaisantes, met également en évidence le type de structures géométriques dont un réseau de neurones a besoin pour être entraîné efficacement. Deuxièmement, la base de données d'apprentissage proposée peut être modifiée en fonction des dispositifs d'acquisition spécifiques et en particulier de leur réponse impulsionnelle, de leur gamme dynamique, des caractéristiques du bruit qui les affecte, etc. Cela ouvre la voie à des schémas d'apprentissage flexibles, génériques et relativement légers.

II. ETAT DE L'ART

A priori de régularité et restauration d'image. Une approche classique de la restauration d'images consiste à supposer une forme close de la distribution des images. Étant donné une observation, on cherche la meilleure explication selon le modèle. Selon la méthode utilisée, la meilleure explication peut être le maximum a posteriori ou le minimiseur du risque quadratique.

Parmi les modèles qui s'inscrivent dans ce cadre général, citons le modèle de Wiener, pour lequel la distribution des images est supposée gaussienne et invariante à la translation, et la variation totale [1] pour laquelle la log-vraisemblance est la norme l^1 du champ vectoriel du gradient. Dans [4], les auteurs dérivent un algorithme de restauration des signaux en supposant que les signaux ciblés sont bien approximés par une représentation parcimonieuse en ondelettes. Par la suite, un ensemble conséquent de publications a discuté des implications de l'hypothèse de parcimonie et une variété d'algorithmes a été proposée pour tirer parti de cette forme particulière de régularité ([5], [6]).

D'autres modèles fructueux ne se traduisent pas aussi facilement par un algorithme de restauration visant à minimiser un critère. Parmi ceux-ci, l'hypothèse d'autosimilarité suppose que l'image possède des répétitions du même motif ou patch : [2], [7]. Les méthodes dérivées de cette hypothèse de régularité sont des algorithmes qui typiquement font la moyenne de patchs similaires afin de co-débruiter ceux qui sont susceptibles de représenter la même vérité terrain.

Les méthodes d'apprentissage profond pour la restauration cherchent directement à construire une machine (le réseau en-

traîné) qui minimise l’erreur de reconstruction. Typiquement, la fonction de perte est prise comme l’erreur quadratique moyenne entre l’image parfaite et la sortie du réseau. Ici, l’a priori sur les images est remplacé par l’ensemble des données d’entraînement qui est censé transmettre suffisamment d’informations sur la distribution des images ([8], [9]). Les efforts sont consacrés à la conception minutieuse du réseau et à ses hyperparamètres.

À son tour, un tel réseau de débruitage peut être utilisé comme un a priori implicite sur la distribution des images. Cette idée, appelée *plug and play*, consiste à utiliser le réseau de débruitage à la place d’un opérateur proximal lors de l’optimisation itérative d’un modèle variationnel [10]. Une autre approche de l’utilisation d’un réseau comme a priori est présentée dans [11] dans lequel il est montré que l’architecture d’un réseau peut servir de régularisateur.

Modèles de synthèse. Certains a priori statistiques sur les images naturelles ont donné lieu à des modèles génératifs capables de synthétiser des images, notamment pour la synthèse de texture : les modèles gaussiens [12], les modèles basés sur les ondelettes [13], [14], les champs de Markov, le modèle feuilles mortes [15], [16], ainsi que les réseaux de neurones [17].

Dans ce travail, nous étudions l’utilisation de ces modèles génératifs comme moyen d’entraîner des réseaux neuronaux de restauration. Nous nous tournons vers le modèle feuilles mortes, en raison de sa simplicité, de son nombre limité de paramètres et de sa capacité à générer des images complexes avec des détails à toutes les échelles.

L’utilisation de modèles synthétiques pour entraîner des réseaux de neurones n’est pas nouvelle et a été largement utilisée pour des tâches d’analyse d’images telles que la segmentation, la reconnaissance ou la détection, voir par exemple [18]. À notre connaissance, aucune stratégie de ce type n’a été utilisée pour des tâches de restauration.

III. GÉNÉRATION D’IMAGES FEUILLES MORTES

A. Le modèle feuilles mortes continu

Le modèle feuilles mortes est un champ aléatoire obtenu par la superposition séquentielle de formes aléatoires. Ce modèle est défini (voir [19]) par un ensemble aléatoire de positions, de temps d’arrivée et de formes $\{(x_i, t_i, X_i)_{i \in \mathbb{N}}$, où $\mathcal{P} = \sum \delta_{x_i, t_i}$ est un processus de Poisson dans $\mathbb{R}^2 \times (-\infty, 0]$ et les X_i sont des ensembles aléatoires de \mathbb{R}^2 indépendants de \mathcal{P} . On appelle feuilles les ensembles $x_i + X_i$, et pour chaque feuille, sa partie visible est définie par

$$V_i = (x_i + X_i) \setminus \bigcup_{t_j \in (t_i, 0)} (x_j + X_j),$$

c’est à dire que la partie visible de la feuille (x_i, t_i, X_i) est obtenue en soustrayant à cette feuille toutes les feuilles d’indice temporel supérieur à t_i (c’est à dire les feuilles tombées après elle). Le modèle feuilles morte est ensuite défini comme l’ensemble des parties visibles. Une image de ce modèle peut être obtenue en attribuant une couleur aléatoire à chaque partie visible.

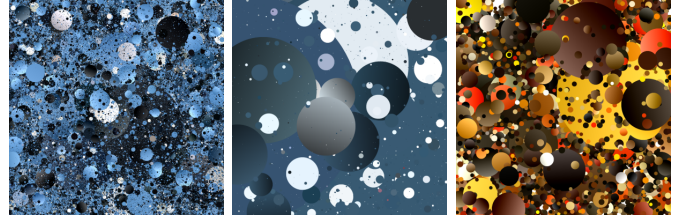


Fig. 1: Exemples d’images feuilles mortes générées avec des paramètres différents. De gauche à droite : $(\alpha = 3, r_{min} = 1, r_{max} = 2000)$, $(\alpha = 2, r_{min} = 1, r_{max} = 2000)$, $(\alpha = 3, r_{min} = 20, r_{max} = 2000)$

Il a été démontré qu’un modèle feuilles mortes où la taille des feuilles est munie de propriétés d’échelle reproduit d’importantes propriétés statistiques des images naturelles [15], [16]. Ces modèles sont obtenus en considérant des feuilles aléatoires $R.X$, où X est une forme donnée et R est une variable aléatoire réelle de densité $f(r) = C.r^{-\alpha}$, où C est une constante de normalisation. Le cas $\alpha = 3$ correspond à un modèle invariant par changement d’échelle [16]. Afin que ces modèles soient bien définis, on contraint R à un intervalle $[r_{min}, r_{max}]$. Le modèle dépend donc de trois paramètres : r_{min} , r_{max} et α . Ce modèle d’images est particulièrement attrayant pour les images naturelles, car il incorpore deux de leurs propriétés fondamentales : la non gaussianité (due à la présence de bords) et les propriétés d’échelles. Comme ce modèle contient des détails et des bords à chaque échelle, il a été proposé comme outil d’évaluation de la capacités des appareils imageurs à respecter les textures [20], [21]. Il a été récemment retenu dans une norme ISO pour l’évaluation de la qualité de ces appareils.

Pour générer des images à partir de ce modèle, nous utilisons l’algorithme proposé et décrit en détails dans [22]. Le modèle choisi génère des disques de rayons aléatoires dont les couleurs sont tirées selon un histogramme d’image naturelle, afin d’éviter les couleurs aberrantes. L’image est ensuite sous-échantillonnée afin d’avoir des bords plus doux, et des détails sub-pixelliques. Des exemples de telles images sont visibles Figure 1 .

IV. RÉSULTATS EXPÉRIMENTAUX

Dans cette partie expérimentale, nous présenterons tout d’abord la base de donnée d’images synthétiques. Ensuite, nous analyserons et comparerons les performances d’entraînement obtenues à partir des images synthétiques à celles obtenues à partir d’images naturelles.

Jeu de données Feuilles Mortes. Afin de prendre en compte à la fois les zones homogènes et les micro-textures, nous construisons un jeu de données généré avec $r_{min} = 1$ ou $r_{min} = 16$ avec à chaque fois $\alpha = 3.0$ et $r_{max} = 2000$. Les micro-textures étant plus difficiles à débruiter, on choisit de générer deux tiers d’images avec $r_{min} = 1$ et un tiers avec $r_{min} = 16$. La distribution des couleurs est donnée par les histogrammes des images de la base d’images naturelles Waterloo [23], pour éviter les couleurs aberrantes qui seraient obtenues si l’on échantillonnait uniformément le cube RVB. Enfin, nous appliquons un flou gaussien à un dixième des images avec un écart-type tiré uniformément entre 1 et 3.

Ceci permet de tenir compte des zones floues des images naturelles, en particulier dues à la profondeur de champ limitée des appareils imageurs. En ajoutant ce simple modèle de flou, nous visons à mieux restaurer ces parties des images naturelles.

Présentation du réseau. Afin d'évaluer la capacité des données synthétiques proposées à entraîner avec succès un réseau de débruitage, nous considérons le réseau FFDNet. Il s'agit d'un CNN de débruitage d'images à l'état de l'art, qui a été introduit par Zhang et al. [9]. Sa principale spécificité repose sur la première couche du réseau : pour augmenter le champ réceptif et gérer une large gamme de niveaux de bruit, l'image est divisée en quatre sous-images qui sont concaténées à une carte de bruit indiquant l'écart-type local du bruit. Ce tenseur est ensuite passé à un réseau plus classique de couches convolutives normalisées par batches, dont l'architecture est similaire à celle du DNCNN. Il produit ensuite quatre sous-images débruitées, qui sont réassemblées pour créer l'image débruitée finale.

Entraînement de FFDNET. Afin de comparer équitablement les différents entraînements, nous utilisons à chaque fois le même algorithme d'optimisation. Il consiste en 80 époques avec l'optimiseur Adam et la fonction de perte L^2 , en commençant avec un taux d'apprentissage de 10^{-3} . On fait décroître ce taux d'un facteur 10 à l'époque 50, puis d'un facteur 100 à l'époque 60. Pour chaque apprentissage, nous avons utilisé 350k (50, 50, 3) patches, extraits soit du jeu de données des feuilles mortes, soit du jeu de données des images naturelles, soit d'un mélange des deux. Le jeu de données mixte contient $\frac{1}{3}$ d'images de feuilles mortes, et $\frac{2}{3}$ d'images naturelles. Pour montrer que les propriétés d'échelle sont nécessaires pour modéliser les images naturelles, nous avons également entraîné FFDNet sur des images de feuilles mortes générées à partir de disques d'un rayon fixe de 100. De plus, nous considérons deux entraînements alternatifs à partir de jeux d'images synthétiques plus simples : des images de bruit blanc et des champs aléatoires gaussiens. L'évaluation numérique est effectuée sur 2 ensembles de tests d'images naturelles (CBSD68, Kodak24) et un ensemble de 24 images de feuilles mortes, générées à partir des couleurs de Kodak24. Pour chaque test, nous calculons le PSNR moyen, le SSIM [24] et la métrique PieAPP [25], une métrique perceptuelle récente, optimisée grâce à des annotations humaines.

Résultats de FFDNET. Sur les deux bases de test d'images naturelles et sur la base de test feuilles mortes, nous observons que le modèle entraîné sur les feuilles mortes surpasse de beaucoup les modèles entraînés sur les bases alternatives d'images synthétiques (0,9 dB pour le modèle de champs gaussien et, sans surprise, 11 dB pour le modèle de bruit blanc), voir le tableau I.

Visuellement, le modèle de champ gaussien conduit à des images débruitées qui contiennent encore du bruit et des artefacts de grille, ce qui a un fort impact sur la métrique PieAPP.

Confirmant notre intuition qu'un modèle d'image avec des propriétés d'échelle est nécessaire, le modèle feuilles mortes avec un rayon fixe a tendance à fortement lisser l'image, perdant ainsi les informations de texture. Ceci se traduit par

une perte de 0.65 dB sur les jeux de test d'images naturelles, et de 1,2 dB sur les images test de feuilles mortes, par rapport au modèle entraîné avec notre version des images feuilles mortes.

De manière plus surprenante, ce modèle n'occasionne qu'une perte de performance modeste, de 0,6 dB, par rapport au modèle entraîné classiquement sur des images naturelles. Visuellement, les résultats sont semblables, malgré certaines limitations. En particulier, l'entraînement synthétique ne permet pas une bonne restauration de certaines lignes fines et à faible contraste, et crée parfois des artefacts sous forme de points. Dans d'autres situations, l'entraînement synthétique améliore les résultats, comme on peut le voir sur la figure 2, où la texture du dessus du champignon est assez bien restaurée (sauf les lignes droites), avec une meilleure préservation des détails fins qu'avec le modèle entraîné sur des images naturelles.

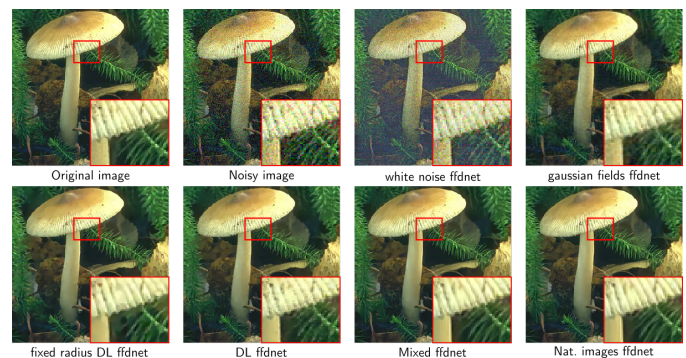


Fig. 2: Comparaisons visuelles des différents modèles de débruitage. De haut en bas de droite à gauche : image non-bruitée, image bruitée avec $\sigma = 25$, FFDNET entraîné sur du bruit blanc, sur des champs gaussiens, des images feuilles mortes de rayon fixe $r = 100$, sur notre modèle feuilles mortes, sur une base mixte, sur des images naturelles.

Un autre résultat est particulièrement intéressant dans un contexte applicatif. L'entraînement sur un mélange d'images de feuilles mortes et d'images naturelles n'affecte pas le résultat du modèle de débruitage sur les bases de test d'images naturelles, la différence de PSNR étant inférieure à 0.02dB. Visuellement, les résultats sont presque identiques, avec un léger avantage pour le modèle entraîné mixte sur les zones de texture. Sur l'ensemble de test des feuilles mortes, le modèle entraîné mixte surpasse le modèle entraîné des images naturelles de 0,25 dB. Ce résultat suggère que l'optimisation conjointe de la réponse à ce type de bases mixtes a la capacité d'améliorer certains aspects sur lesquels les dispositifs d'imagerie sont évalués. En effet, l'invariance au changement d'échelle des feuilles mortes est utilisée pour évaluer la capacité des appareils imageurs, et en particulier des smartphones, à préserver les zones de texture ; la mire correspondante, invariable à l'échelle, est à la base d'une norme ISO (ISO/TS 19567-2:2019).

Dans [22], nous montrons que ces mêmes images synthétiques peuvent être utilisées pour l'apprentissage de tâches de super-résolution mono-image. Ces expériences confirment l'adéquation aux tâches de restauration d'images du modèle feuilles mortes muni de propriétés d'échelle et d'une distribution des couleurs appropriée.

TABLE I: Résultats numériques des différents entraînements de FFDNet. On a testé les modèles sur deux bases standards pour le débruitage (CBSD68 et Kodak24), et sur une base synthétique, pour deux niveaux de bruits. Chaque case contient un triplet PSNR/SSIM/PieAPP. Les meilleurs résultats sont en bleu, les seconds en rouge.

σ	Dataset	CBSD68	Kodak24	Dead leaves testset
$\sigma = 25$	White Noise	19.52/0.416/2.386	19.68/0.365/2.502	20.36/0.607/2.043
	Gaussian field	29.63/0.845/1.402	30.24/0.835/1.471	26.23/0.826/1.254
	DL $r = 100$	29.56/0.820/1.218	30.49/0.819/1.024	26.13/0.799/1.263
	Dead leaves	30.58/0.867/0.711	31.27/0.859/0.739	27.46/0.865/0.573
	Mix	31.07/0.881/0.639	31.98/0.876/0.603	27.33/0.860/0.567
	Natural Images	31.09/0.882/0.629	32.00/0.878/0.599	27.05/0.851/0.576
$\sigma = 50$	White Noise	15.58/0.247/4.682	15.71/0.209/4.785	16.24/0.387/2.932
	Gaussian field	26.68/0.738/2.203	27.41/0.737/2.353	23.31/0.694/2.158
	DL $r = 100$	26.85/0.720/1.563	27.91/0.739/1.314	23.24/0.654/2.005
	Dead leaves	27.40/0.762/1.088	28.21/0.765/1.154	24.21/0.737/1.020
	Mix	27.86/0.782/0.997	28.86/0.789/0.985	24.12/0.732/1.015
	Natural Images	27.87/0.786/0.991	28.89/0.792/0.978	23.90/0.722/1.053

V. CONCLUSIONS ET PERSPECTIVES

A notre connaissance, ce travail est le premier où l'on entraîne un réseau de restauration d'images sur des images synthétiques. Nous avons tout d'abord présenté le modèle feuilles mortes ainsi que les choix des paramètres de génération pour notre base synthétique. Tant pour le débruitage que pour la super-résolution (voir [22]), les modèles entraînés sur la base de données feuilles mortes ont des performances étonnamment proches de ceux entraînés sur des images naturelles. En mélangeant des images naturelles et synthétiques dans l'entraînement, les résultats atteignent des performances équivalentes à celles du modèle entraîné sur des images naturelles uniquement. Ces deux résultats indiquent que le modèle feuilles mortes avec des propriétés d'échelle est un bon candidat pour remplacer les images naturelles pour l'entraînement, avec seulement quelques paramètres. En effet, la géométrie du modèle ne dépend que de trois paramètres : α, r_{min}, r_{max} . Même si les paramètres de couleur sont encore relativement nombreux, une piste alternative est de recourir à des échantillonnages simples d'espaces de couleur perceptuels afin d'éviter les couleurs artificielles. Une autre perspective est de compléter notre jeu de données, par exemple avec des patches sinusoïdaux, pour mieux restituer les patches oscillants et les lignes droites. Enfin, nous pensons, comme nous l'avons déjà expliqué, qu'un ensemble de données synthétiques peut être un moyen simple d'éviter les campagnes d'acquisition relativement lourdes [3] impliquées par chaque nouvelle modalité d'acquisition et nous prévoyons d'approfondir cette piste.

REFERENCES

- [1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, 1992.
- [2] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, 2005.
- [3] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [4] D. L. Donoho, I. M. Johnstone *et al.*, "Minimax estimation via wavelet shrinkage," *The annals of Statistics*, 1998.
- [5] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, 2005.
- [6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, 2006.
- [7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," in *Image Processing: Algorithms and Systems VI*, International Society for Optics and Photonics. SPIE, 2008.
- [8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, 2017.
- [9] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, 2018.
- [10] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Processing Letters*, 2017.
- [11] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.
- [12] B. Galerne, Y. Gousseau, and J.-M. Morel, "Micro-texture synthesis by phase randomization," *Image Processing On Line*, vol. 1, pp. 213–237, 2011.
- [13] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of the 22nd annual conference on computer graphics and interactive techniques*, 1995.
- [14] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International journal of computer vision*, 2000.
- [15] L. Alvarez, Y. Gousseau, and J.-M. Morel, "The size of objects in natural and artificial images," in *Advances in Imaging and Electron Physics*. Elsevier, 1999.
- [16] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *International Journal of Computer Vision*, 2001.
- [17] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," 2015.
- [18] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.
- [19] C. Bordenave, Y. Gousseau, and F. Roueff, "The dead leaves model: a general tessellation modeling occlusion," *Advances in applied probability*, 2006.
- [20] F. Cao, F. Guichard, and H. Hornung, "Measuring texture sharpness of a digital camera," in *Digital Photography V*, vol. 7250. International Society for Optics and Photonics, 2009, p. 72500H.
- [21] —, "Dead leaves model for measuring texture quality on a digital camera," in *Digital Photography VI*. International Society for Optics and Photonics.
- [22] R. Achddou, Y. Gousseau, and S. Ladjal, "Synthetic images as a regularity prior for image restoration neural networks," in *Scale Space and Variational Methods in Computer Vision*. Springer International Publishing, 2021.
- [23] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, 2016.
- [24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, 2004.
- [25] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "Pieapp: Perceptual image-error assessment through pairwise preference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.