

Synthèse de bruit par modèle de diffusion pour la restauration d'images en basse luminosité

Liyang LU¹ Raphael ACHDDOU¹ Sabine SÜSSTRUNK¹

¹IC, IVRL, École Polytechnique Fédérale de Lausanne, BC 328 (Bâtiment BC) Station 14, 1015 Lausanne

Résumé – Obtenir des images de bonne qualité en basse luminosité est difficile en raison d'un rapport signal-à-bruit très bas. Les modèles de bruit usuels, comme le modèle Gaussien, sont insuffisants dans ces conditions. L'entraînement de réseaux neuronaux de restauration nécessite donc des données réelles appariées ou un modèle précis du bruit en basse luminosité. Cet article explore l'utilisation de modèles de diffusion génératifs pour résoudre ce problème. Après avoir adapté ces modèles pour une modélisation précise du bruit, nous démontrons que notre méthode permet de créer des bases de données de tailles arbitraires, améliorant ainsi les résultats des réseaux neuronaux de débruitage.

Abstract – Obtaining high-quality images in low light is challenging due to an extremely low signal-to-noise ratio. Standard noise models, such as the Gaussian model, are insufficient under these conditions. Therefore, training neural networks for image restoration requires either real paired data or an accurate low-light noise model. This article explores the use of generative diffusion models to address this issue. After adapting these models for precise noise modeling, we demonstrate that our method can create arbitrarily sized datasets, improving the results of denoising neural networks.

1 Introduction

Obtenir des photographies sans bruit dans des conditions de faible luminosité constitue un défi majeur. L'augmentation du temps d'exposition améliore la luminosité dans les scènes statiques, mais crée du flou dans les scènes dynamiques. De même, l'augmentation du gain (ISO) amplifie la luminosité mais également le bruit, nécessitant donc des algorithmes de débruitage performants. Or, les méthodes de débruitage par apprentissage profond nécessitent de grandes quantités de données et la collecte d'images bruitées-propres appariées est fastidieuse, nécessitant une caméra fixe et l'absence d'objets en mouvement, ce qui limite les données aux scènes statiques.

Pour pallier cette difficulté, des techniques de synthèse de bruit, incluant des modèles physiques ou des modèles génératifs neuronaux ont été développées. Les modèles physiques, utilisant des variations du modèle de bruit Poisson-Gaussien [7], peinent à reproduire les distributions de bruit réelles. Les méthodes d'apprentissage profond, telles que les GAN et les flux de normalisation [1, 8], rencontrent des problèmes de stabilité d'entraînement et de gestion des distributions multimodales.

Récemment, les modèles de diffusion [5] ont permis des avancées majeures dans la modélisation de distributions de données complexes, comme la distribution des images naturelles. Ces modèles s'avèrent donc être un candidat naturel pour la modélisation du bruit en basse luminosité. Cet article explore cette perspective, en proposant un algorithme de diffusion avec différentes adaptations cruciales : (1) un réseau neuronal à deux branches pour modéliser le bruit de lecture et le bruit photonique, (2) un encodage positionnel pour le bruit spatialement variable, (3) et un schéma de diffusion optimisé. Ces améliorations permettent de modéliser le bruit très précisément et conduisent naturellement à des performances de débruitage supérieures à l'état de l'art, comme illustré dans la Figure 1.



FIGURE 1 : Résultats de réseaux neuronaux de débruitages entraînés avec différents générateurs de bruit synthétiques vs données réelles appariées. La qualité des résultats dépend fortement de celle du modèle de bruit choisi. Notre méthode produit les meilleurs résultats.

2 État de l'art

La restauration d'images en basse luminosité revient à un problème de débruitage avec des rapports signal-à-bruits extrêmement bas. Le débruitage étant un problème central du traitement d'image, il existe de nombreuses méthodes, qu'elles soient basées sur des a priori mathématiques [3, 6], ou sur des algorithmes par apprentissage profond [9]. La plupart de ces modèles reposent sur l'hypothèse d'un bruit blanc Gaussien centré. Cette approximation est malheureusement insuffisante en basse luminosité. Chen et al [2] ont donc proposé de collecter un ensemble de paires d'images (longue/courte exposition) comme base d'apprentissage pour un réseau neuronal de restauration de type UNet, obtenant des résultats bien meilleurs que toutes les méthodes basées sur le modèle Gaussien.

Cependant, l'acquisition de ce genre de base est fastidieuse, et sa taille limitée ne permet pas d'obtenir des performances de restauration optimales. Afin de pallier ce problème, la synthèse de bruit réaliste a suscité beaucoup d'intérêt. Une première possibilité consiste à modéliser le bruit de capteur avec des

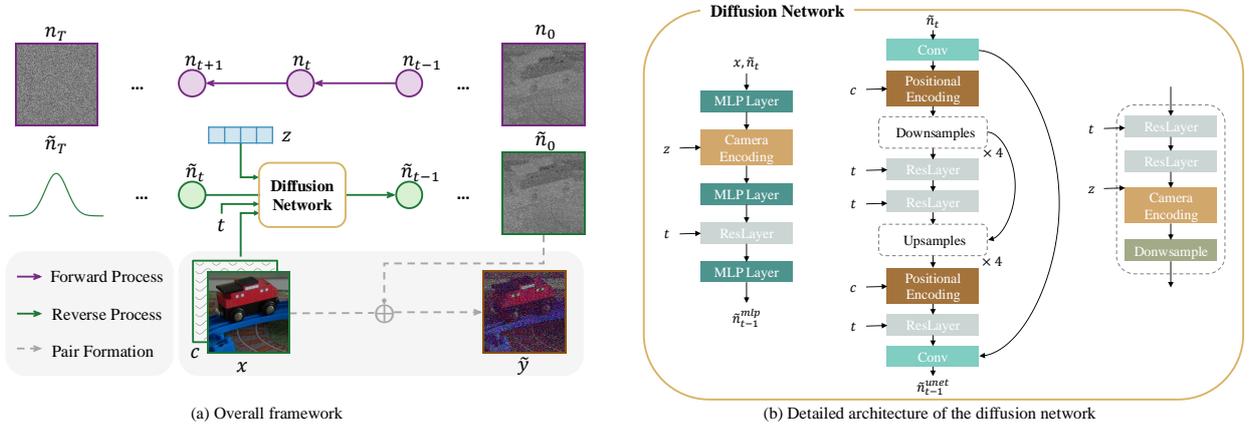


FIGURE 2 : Présentation de la méthode de génération du bruit basse lumière par méthode de diffusion. La première figure résume l’algorithme génération, et la seconde l’architecture du réseau de diffusion.

modèles physiques plus précis que le modèle Gaussien. Wei et al [7] ont notamment proposé un modèle introduisant différentes composantes du bruit de capteur, dont les paramètres sont optimisés sur une base d’images bruitées-propres appariées. D’autres méthodes basées sur une séparation du bruit de lecture et du bruit photonique, ne génèrent que le bruit photonique avec un modèle de Poisson et échantillonnent une base de données de réalisation du bruit de lecture obtenues dans l’absence totale de lumière [4].

Une autre classe de méthode s’est ensuite dégagée suite aux progrès des méthodes d’intelligence artificielle générative. Les réseaux adverses (GANs) [8] et les méthodes par flots de normalisation [1] ont obtenus des résultats satisfaisants. Cependant, ces méthodes sont limitées par leurs instabilités et leurs faibles capacités à modéliser des distributions complexes. Cet article présente la première tentative fructueuse d’adaptation des modèles de diffusion à cette tâche.

3 Diffusion pour la synthèse de bruit

L’application directe des modèles de diffusion probabilistes par débruitage (DDPM [5]) ne permet pas d’obtenir de bons résultats. On présente ici les adaptations nécessaires à la modélisation précise du bruit en basse lumière. La première consiste à modifier l’architecture du réseau de génération pour mieux prendre en compte le bruit fixe ou *fixed-pattern noise* (FPN). La seconde consiste à modifier le schéma de bruit lors de la génération pour mieux préserver la variance du bruit généré.

3.1 Cadre général

Le cadre général est illustré dans la Figure 2 (a). L’entrée initiale pour le processus direct du modèle de diffusion est le bruit en faible luminosité n_0 , obtenu en soustrayant l’image propre de l’image bruitée. Un bruit Gaussien est progressivement ajouté à n_0 jusqu’à obtenir une distribution normale au temps $T = 1000$. Le processus inverse utilise un réseau pour prédire ce bruit, ramenant \tilde{n}_T à \tilde{n}_0 . Ce processus permet d’échantillonner du bruit réel à partir d’une variable aléatoire Gaussienne.

Le processus est conditionné par l’image sans bruit et les paramètres de la caméra pour modéliser le bruit dépendant du signal. Les positions des pixels sont encodées dans le réseau

pour capturer les composantes de bruit spatialement corrélées, comme le bruit de motif fixe (FPN). Le processus inverse est donné par $\tilde{n}_t = f_\theta(\tilde{n}_{t-1}, t, \mathbf{x}, \mathbf{c}, \mathbf{z})$, où t est l’étape de diffusion, \mathbf{x} est l’image sans bruit, \mathbf{c} représente les coordonnées des pixels, et \mathbf{z} inclut les paramètres de la caméra.

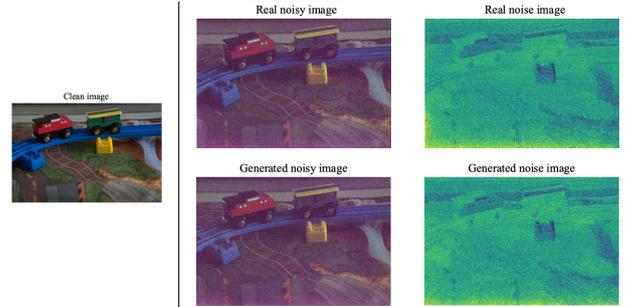


FIGURE 3 : Exemple d’une paire d’image sans bruit/bruitée de la base SID, et d’un résultat de génération obtenu avec notre méthode. Les cartes de bruits correspondantes sont très proches.

3.2 Adaptations architecturales

Architecture à deux branches On propose une architecture du réseau de diffusion à deux branches, dont le fonctionnement est présenté en détail dans la Figure 2. La première est un Perceptron Multi Couches (MLP), traitant indépendamment chaque pixel. La seconde est un réseau convolutif de type UNet. Ces choix architecturaux tiennent compte de la nature distincte des différentes composantes du bruit. Par exemple, le bruit photonique ne dépend que de la valeur du pixel sans bruit, faisant du réseau MLP un choix approprié. En revanche, les composantes de bruit localement corrélées sont mieux capturées par l’architecture UNet, qui est un opérateur local par construction. L’architecture du réseau intègre aussi un encodage positionnel pour le réseau UNet et un encodage des paramètres d’acquisition.

Encodage positionnel La distribution du bruit en basse lumière en chaque pixel dépend fortement de leur position dans le capteur. Comme illustré dans la Figure 3, on observe un signal basse fréquence non uniforme sur la totalité de l’image.

Afin de modéliser ce genre de signal, nous intégrons un encodage positionnel dans le réseau de diffusion. Nous utili-

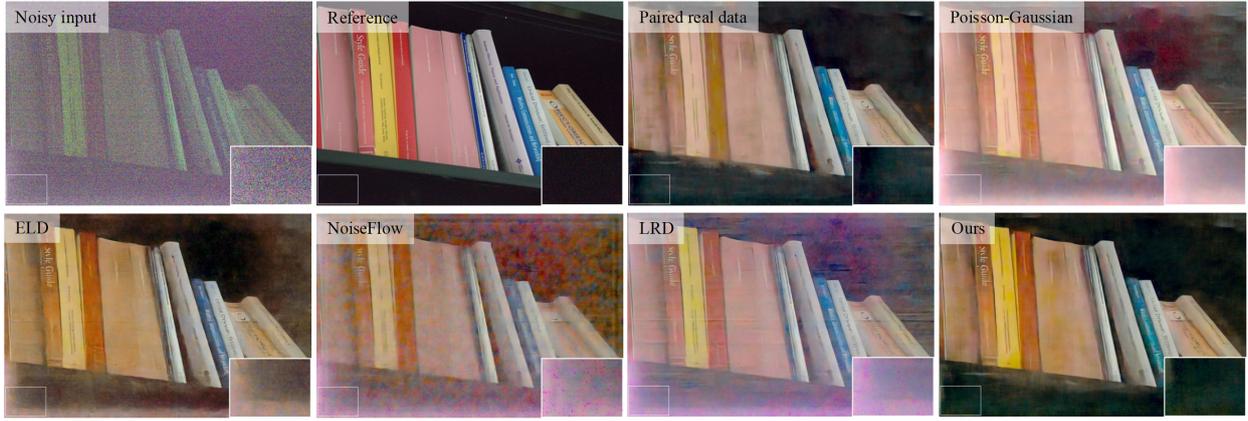


FIGURE 4 : Comparaison des résultats obtenus à partir de différents entraînement d’un réseau de débruitage UNet en fonction du générateur de bruit choisi. Notre méthode permet de corriger les composantes basse fréquence et les biais du bruit, et de supprimer le bruit photonique.

sons un codage positionnel sinusoïdal produisant une carte de position \mathbf{c} , transformée en une carte d’échelle et de décalage par une convolution 1×1 . Cette information positionnelle est intégrée en appliquant la transformation suivante point à point aux représentations cachées du réseau : $\tilde{\mathbf{f}} = \mathbf{f} \odot (1 + \mathbf{s}) + \mathbf{r}$.

Encodage des paramètres d’acquisition Les paramètres d’acquisition (comme l’ISO et le temps d’exposition) influencent la distribution du bruit. C’est pourquoi nous encodons ces informations, en créant une représentation neuronale (“embedding”) des paramètres d’acquisition, optimisé durant l’apprentissage. L’embedding de chaque paramètre d’acquisition est intégré dans les représentations cachées du réseau en utilisant des mécanismes d’attention croisée.

3.3 Schéma de diffusion

Modéliser précisément la variance est un critère de qualité important de la génération de bruit en basse luminosité. En effet, le moindre écart nuirait aux résultats de débruitage. Nous montrons que le choix du schéma de diffusion joue un rôle crucial pour ce problème. Il est intéressant de constater que le schéma classique de diffusion en cosinus réduit considérablement la variance du bruit généré. Nous attribuons cet écueil à l’entraînement des modèles de diffusion par débruitage qui minimise l’erreur quadratique moyenne. L’espérance de l’estimateur associé s’approche de la moyenne de toutes les solutions possibles. Plus la variance du bruit Gaussien est forte, plus le problème est mal posé, et plus la solution est lisse.

Supposons qu’une image bruitée v suit la relation $v = u + \epsilon$, où v est l’image originale et $\epsilon \sim \mathcal{N}(0, \sigma_1^2)$ correspond au bruit ajouté dans le processus direct de diffusion. Ici, la distribution objectif U est la distribution du bruit en basse luminosité. Pour simplifier, nous supposons que U suit une distribution Gaussienne $\mathcal{N}(\lambda, \sigma_2^2)$.

Proposition 1. *Lorsqu’un débruiteur est entraîné à minimiser l’erreur quadratique moyenne, la solution optimale est $f_{MMSE}(v) = \mathbb{E}[u|v]$, et la variance de l’image produite vaut $\frac{\sigma_2^4}{\sigma_2^2 + \sigma_1^2}$.*

Ce résultat suggère que plus la variance du bruit Gaussien ajouté à la dernière étape de diffusion est grande, plus la variance du bruit basse lumière générée sera réduite. Nous avons observé ce phénomène pour plusieurs types de schéma de diffusion, comme le schéma en cosinus, et des schémas logistiques

aux accroissements forts. Plus l’accroissement diminue, plus ce phénomène s’atténue. Nous choisissons donc de nous orienter vers ce type de schéma dans la suite de nos expériences.

4 Validation Expérimentale

Afin d’évaluer la précision de notre modèle, deux approches sont possibles. La première consiste à utiliser le générateur pour entraîner un réseau de débruitage en basse luminosité et d’en évaluer les performances. La seconde consiste à mesurer des statistiques du bruit généré et à les comparer à celles du bruit réel. On se propose de mesurer la moyenne et la variance du bruit pour chaque intensité, et à mesurer la divergence de Kullback-Leibler entre le bruit généré et le bruit réel.

4.1 Expériences de débruitage

Afin de se comparer aux autres méthodes standards de génération de bruit, on utilise un réseau UNet standard pour le débruitage. On génère des images de bruit de taille (512×512) pour entraîner le modèle. L’entraînement suit les mêmes principes que la méthode SID, soit une minimisation de la fonction de perte L_1 avec ADAM.

On rapporte les résultats de débruitages évalués sur la base SID[2] et la base ELD[7] dans la Table 1. Notre méthode permet d’obtenir les meilleurs résultats avec des écarts importants pour toutes les configurations, exceptés le rapport d’exposition $\times 200$ de la base ELD. Ceci est dû au fait que ce paramètre d’acquisition n’existe pas dans la base SID sur laquelle notre modèle de diffusion a été entraîné.

Visuellement, l’exemple présenté dans la Figure 4 corrobore les résultats numériques. Les couleurs sont plus vives et plus fidèlement restaurées avec notre modèle. De plus, les composantes basses fréquences du bruit FPN sont corrigées efficacement, là où les autres méthodes échouent.

4.2 Étude d’ablation

Outre les performances de débruitage, il est difficile de se comparer aux autres modèles de l’état de l’art, leurs codes de génération de bruit n’étant pas disponibles. Néanmoins, l’étude statistique du bruit généré nous permet de mieux comprendre quels sont les composants importants de notre modèle. On étudie plusieurs configurations d’ablation : une architecture UNet mono-branche, un modèle sans encodage positionnel, et

Dataset	Ratio	Bruit réel	Modèles physiques		Modèles neuronaux		
		images apparées	Poisson-Gaussien	ELD [7]	NoiseFlow [1]	LRD [8]	Diffusion
SID	×100	42.95 / 0.958	41.05 / 0.936	<u>41.95 / 0.953</u>	40.21 / 0.925	40.97 / 0.930	43.30 / 0.958
	×250	40.27 / 0.943	36.63 / 0.885	<u>39.44 / 0.931</u>	36.15 / 0.870	37.00 / 0.875	40.53 / 0.944
	×300	37.32 / 0.928	33.34 / 0.811	<u>36.36 / 0.911</u>	33.27 / 0.803	33.74 / 0.803	37.68 / 0.928
ELD	×100	45.52 / 0.977	44.02 / 0.931	<u>45.45 / 0.975</u>	43.31 / 0.941	44.56 / 0.946	<u>45.79 / 0.972</u>
	×200	41.70 / 0.912	41.00 / 0.880	43.43 / 0.954	40.26 / 0.885	41.71 / 0.899	<u>42.25 / 0.924</u>

TABLE 1 : Évaluation numérique sur SID et ELD en termes de PSNR et SSIM pour divers ratios d'exposition. Les meilleurs résultats sont en gras, et les deuxièmes sont soulignés.

un algorithme d'échantillonnage déterministe rapide (DDIM). Les résultats de l'étude d'ablation, présentés dans la Figure 5, suggèrent différentes remarques. (1) Le réseau UNet seul ne peut modéliser correctement la moyenne du bruit, car on voit des artefacts de damier sur l'image générée. (2) Supprimer l'encodage positionnel ne permet pas de modéliser le bruit FPN, et en particulier sa composante basse fréquence. (3) Utiliser un algorithme d'échantillonnage plus rapide, et donc avec un bruit plus fort à la dernière étape de diffusion mène à une variance plus faible du bruit généré, ce qui se traduit aussi par une divergence KL plus importante.

5 Conclusion

Cet article présente la première tentative fructueuse d'application des modèles de diffusion à la synthèse de bruit en basse luminosité. Les différentes adaptations du modèle de diffusion DDPM, concernant l'architecture, le conditionnement, et le schéma de bruit, permettent d'obtenir des performances de débruitage en basse lumière supérieures à l'état de l'art.

Références

- [1] Abdelrahman ABDELHAMED, Marcus A BRUBAKER et Michael S BROWN : Noise flow : Noise modeling with conditional normalizing flows. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3165–3173, 2019.
- [2] Chen CHEN, Qifeng CHEN, Jia XU et Vladlen KOLTUN : Learning to see in the dark. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.
- [3] Kostadin DABOV, Alessandro FOI, Vladimir KATKOVNIK et Karen EGIAZARIAN : Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.
- [4] Hansen FENG, Lizhi WANG, Yuzhi WANG, Haoqiang FAN et Hua HUANG : Learnability enhancement for low-light raw image denoising : a data perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):370–387, 2023.
- [5] Jonathan HO, Ajay JAIN et Pieter ABBEEL : Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- [6] Leonid I RUDIN, Stanley OSHER et Emad FATEMI : Non-linear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1-4):259–268, 1992.
- [7] Kaixuan WEI, Ying FU, Yinqiang ZHENG et Jiaolong YANG : Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8520–8537, 2021.
- [8] Feng ZHANG, Bin XU, Zhiqiang LI, Xinran LIU, Qingbo LU, Changxin GAO et Nong SANG : Towards general low-light raw noise synthesis and modeling. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10820–10830, 2023.
- [9] Kai ZHANG, Wangmeng ZUO, Yunjin CHEN, Deyu MENG et Lei ZHANG : Beyond a gaussian denoiser : Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

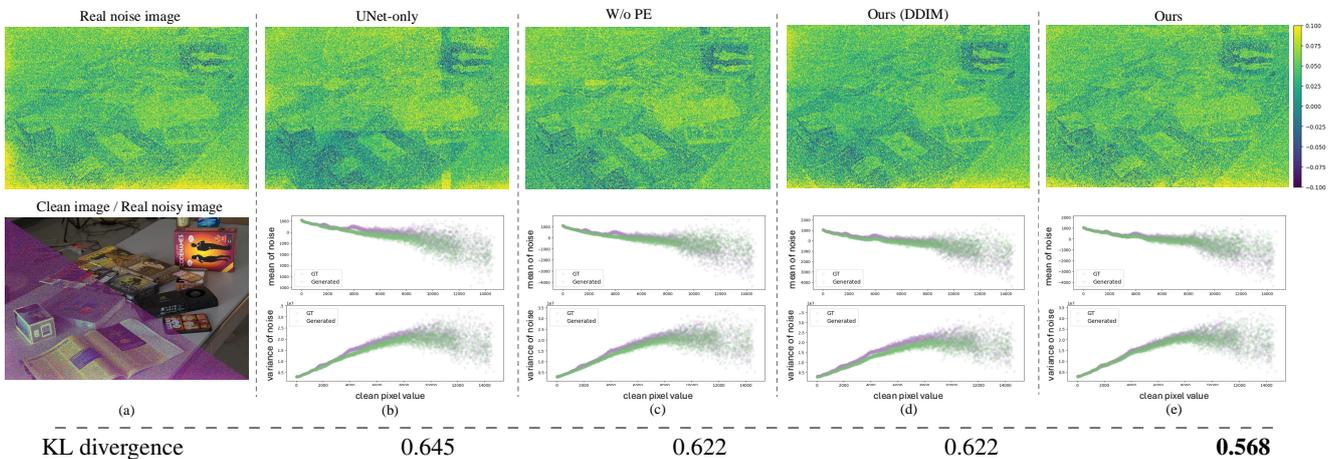


FIGURE 5 : Résultats statistiques de l'étude d'ablation. De haut en bas : Image de bruit générées / Courbes de variances et moyennes / KLD