

Approche couplée VAE et interpolateur pour l'émulation rapide d'images hyperspectrales

Chedly BEN AZIZI Claire GUILLOTEAU Gilles ROUSSEL Matthieu PUIGT

Univ. Littoral Côte d'Opale, LISIC – UR 4491, 62219 Longuenesse, France

Résumé – La génération de données hyperspectrales de haute fidélité est un élément essentiel des missions d'observation, mais elle implique traditionnellement des simulations numériques coûteuses. Ce défi peut être relevé en construisant un modèle statistique approchant la distribution des données simulées. Ce processus est appelé émulation. Cet article présente une nouvelle approche d'émulation couplant un autoencodeur variationnel (VAE) et un interpolateur pour une émulation rapide d'images hyperspectrales. Cette approche permet d'émuler l'image hyperspectrale à partir des paramètres biophysiques de manière efficace et précise. Nous en dérivons deux méthodes : une méthode d'émulation de cubes hyperspectraux et une méthode d'émulation spectre-à-spectre. Ces méthodes sont évaluées sur les images hyperspectrales Sentinel-3 et montrent des améliorations significatives en termes de précision et d'efficacité par rapport aux méthodes existantes.

Abstract – Generating high-fidelity hyperspectral data is an essential part of observation missions, but traditionally involves costly numerical simulations. This challenge can be met by building a statistical model that approximates the distribution of the simulated data. This process is called emulation. This paper presents a new emulation approach coupling a variational autoencoder (VAE) and an interpolator for fast emulation of hyperspectral images. This approach enables the hyperspectral image to be emulated from the biophysical parameters efficiently and accurately. We derive two methods from this framework: a hyperspectral cube emulation method and a spectrum-to-spectrum emulation method. These methods are evaluated on Sentinel-3 hyperspectral images and shows significant improvements in terms of accuracy and efficiency compared with existing methods.

1 Introduction

L'imagerie hyperspectrale combine les avantages de l'imagerie et de la spectroscopie. Elle fournit ainsi une description spatiale 2D et spectrale 1D d'une scène, produisant un cube de données tridimensionnel à haute résolution spectrale, appelé image hyperspectrale (HS). Ses applications s'étendent à un large éventail de domaines, dont l'astrophysique, la surveillance agricole ou la télédétection [2]. La génération de données de haute fidélité constitue un enjeu majeur pour le succès des missions d'observation hyperspectrale, que ce soit pour le développement anticipé d'algorithmes de traitement de données dédiés ou pour la conception de futures missions. Dans le cas particulier de la télédétection, des simulateurs capables de synthétiser des images satellites à partir de paramètres biophysiques ont été développés [4]. Cependant, modéliser des phénomènes physiques non-linéaires et complexes nécessite généralement des simulations numériques coûteuses.

Le développement récent et rapide des techniques d'apprentissage automatique a conduit à la création de modèles statistiques capables d'approcher le comportement de simulateurs avec un haut degré de précision pour une fraction du coût de calcul. Ces modèles, appelés émulateurs, sont en pratique construits à l'aide d'un ensemble de paires entrée-sortie d'entraînement. Cet ensemble d'entraînement est ensuite utilisé pour apprendre les paramètres d'un modèle probabiliste. Les approches pionnières se sont concentrées sur les méthodes statistiques bayésiennes [5] tandis que les travaux plus récents se focalisent sur l'utilisation de modèles d'apprentissage tels que les réseaux de neurones artificiels (RNA) [10].

Les RNA et l'apprentissage profond en général ont été initialement utilisés dans l'imagerie hyperspectrale pour les tâches

de classification. Plus récemment, cependant, les réseaux génératifs profonds (RGP) ont été introduits pour résoudre des tâches telles que le démixage spectral [1]. Les RGP sont des modèles capables d'apprendre la distribution de probabilité sous-jacente aux données d'apprentissage, sans qu'il soit nécessaire d'évaluer explicitement la fonction de densité. Grâce à cette propriété, les RGP permettent de synthétiser des données réalistes en échantillonnant à partir de cette distribution, ce qui les rend adéquats pour l'émulation. L'utilisation des RGP en imagerie hyperspectrale se focalisent principalement sur des tâches de classification et segmentation, super-resolution, etc. Plus récemment, des méthodes basées sur les réseaux génératifs antagonistes (RGA) pour la génération d'images de Sentinel 2 (multispectrale) ont été développées[8].

Bien que les méthodes d'émulation d'images HS existantes aient apporté des contributions significatives, elles se concentrent principalement sur la synthèse spectre à spectre, ce qui limite leur capacité à capturer les complexités inhérentes aux informations spatiales et spatio-spectrales dans les données HS. En outre, les émulateurs hyperspectraux basés sur des techniques d'apprentissage automatique commencent seulement à émerger [3]. Dans ce travail, nous proposons une approche cadre à l'émulation d'images HS. Cette approche se déroule en deux phases : une phase de réduction de dimension et une phase d'interpolation. L'originalité du travail réside dans l'utilisation d'un modèle de réseau génératif profond, l'autoencodeur variationnel (VAE) pour la phase de réduction de dimension. La seconde phase consiste à entraîner un interpolateur à faire correspondre un ensemble de paramètres biophysiques à l'espace latent appris. L'émulation d'images HS à partir des paramètres biophysiques est alors l'interpolation de ces derniers dans l'espace latent appris et leur décodage en images. Le pro-

cessus en deux étapes nous permet de faire la distinction entre la synthèse et l'interpolation. Deux méthodes dérivées sont proposées dans cet article : une méthode d'émulation de cubes hyperspectraux et une méthode d'émulation spectre-à-spectre. Les résultats expérimentaux soulignent les améliorations significatives par rapport aux méthodes existantes. L'approche globale d'émulation spatio-spectrale d'images entières présente des performances supérieures.

L'article est organisé de la manière suivante. La section 2 présente l'approche proposée en détaillant le modèle d'émulation en deux étapes. Puis, la section 3 décrit le dispositif expérimental et examine les résultats obtenus. Enfin, la section 4 résume les principales conclusions de l'étude et les perspectives de ce travail.

2 Approche proposée

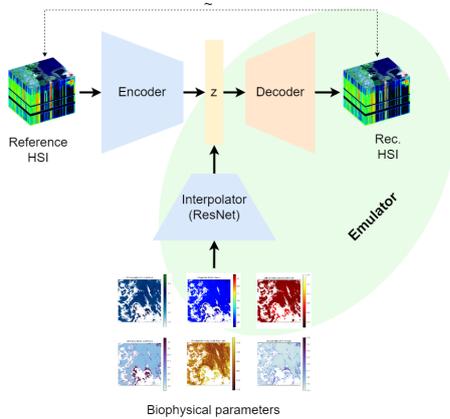


FIGURE 1 – Modèle d'émulation proposé : un VAE est d'abord entraîné pour reproduire des images hyperspectrales en entrée. Ensuite, l'encodeur est écarté et un interpolateur est entraîné pour relier les paramètres biophysiques à l'espace latent appris. L'émulation se fait donc par l'interpolation d'une entrée donnée dans l'espace latent, puis par la génération d'un cube à l'aide du décodeur.

Traditionnellement, l'émulation d'images HS est réalisée en deux étapes. Une première phase de réduction de dimension est appliquée aux données, typiquement via une analyse en composantes principales effectuée sur les spectres de l'image. Une seconde phase prédit les coefficients de projection de nouveaux spectres à l'aide d'un modèle de régression [10]. Cela permet de contourner les limites des modèles de régression lorsqu'ils sont appliqués sur des données de grande dimension telles que les HSI. Nous choisissons donc naturellement ce cadre pour développer notre approche.

La figure 1 décrit les deux étapes de la méthode que nous proposons. Nous choisissons une structure d'auto-encodeur variationnel (VAE) pour l'étape de réduction de dimension. L'VAE est un modèle génératif qui encode une donnée \mathbf{I} en une représentation latente z . Ensuite, le décodeur utilise ce code pour reconstruire l'image initiale, soit :

$$z \sim E(x) = q(z|\mathbf{I}), \hat{\mathbf{I}} \sim D(z) = p(\mathbf{I}|z), \quad (1)$$

où p et q représentent respectivement le *prior* et le *posterior*. Ce choix est guidé par deux avantages principaux de ce type de

modèle. Premièrement, contrairement aux méthodes linéaires, un VAE est capable de modéliser les relations non-linéaires entre les spectres et entre les bandes spectrales. Cette propriété permet de préserver la complexité des signatures spectrales présentes dans les images HS [11]. Deuxièmement, l'approche variationnelle contraint l'espace latent à suivre une distribution *a priori* gaussienne. Cela permet un échantillonnage aléatoire cohérent : tout échantillon de l'espace latent produit une sortie plausible et deux échantillons proches de l'espace latent génèrent des échantillons similaires [7].

Lorsque le VAE est entraîné et que la tâche de reproduction fidèle des HSI est validée, l'encodeur est écarté. Un interpolateur f_ϕ permet ensuite de réaliser la seconde phase en liant les variables biophysiques d'entrée à l'espace latent. f_θ est choisie de sorte à minimiser une fonction coût \mathcal{L} entre les images émules $\hat{\mathbf{I}} = D_\theta(f_\phi(\mathbf{Y}))$ et les images réelles correspondantes \mathbf{I} , soit :

$$\min_{\phi} \mathcal{L}(\theta, \phi, \mathbf{Y}^{(i)}) = \min_{\phi} \|\mathbf{I}^{(i)} - D_\theta(f_\phi(\mathbf{Y}^{(i)}))\|_2. \quad (2)$$

où D_θ représente le générateur de paramètre θ obtenu lors de la première étape et \mathbf{Y} les paramètres biophysiques. Ce module d'interpolation peut prendre différentes formes de modèle comme un processus gaussien, un réseau de neurones ou une méthode hybride.

En outre, cette approche couplée VAE-interpolateur présente deux avantages supplémentaires. Le premier avantage réside dans le fait que l'entraînement et la validation de chaque composante du modèle (VAE pour la réduction de dimension et interpolation) peut être faite de manière séquentielle et indépendante. Cela facilite le réglage des hyperparamètres. Le second avantage est lié à l'adaptabilité du modèle. En effet, toute modification des paramètres biophysiques d'entrée, sans changement des propriétés des images HS à générer, ne nécessite que le ré-entraînement de l'interpolateur, ce qui représente un gain de temps important.

Nous proposons et comparons deux implémentations différentes de la même approche présentée ci-dessus. La première, baptisée HSEmule, émule des cubes entiers [3], permettant la génération des caractéristiques spatio-spectrales. À la fois l'encodeur et l'interpolateur sont des réseaux de type ResNet-18[6]. Le décodeur inverse l'architecture et intègre un sur-échantillonnage suivi d'une convolution, selon la méthode d'Odena *et al.* [9], afin d'atténuer les artefacts de damier résultant des opérations de convolution transposées généralement utilisées pour les décodeurs. Cette approche nécessite un entraînement prolongé de 400 époques. La seconde méthode consiste en une émulation spectre-à-spectre. Plutôt que de générer des cubes complets, cet émulateur produit des spectres. Un cube entier peut ensuite être construit à partir de la génération et la réorganisation d'un certain nombre de spectres. Dans cet objectif, le modèle utilisé pour la phase de réduction de dimension est un réseau à trois couches denses (128, 512 puis 256 neurones). La même architecture est utilisée pour l'encodeur et pour l'interpolateur. Le décodeur est, ici aussi, construit symétriquement en inversant les couches de l'encodeur. Cette configuration a l'avantage d'être entraînée plus rapidement (100 époques). Nous nommons cette méthode p2p-HSEmule. Une implémentation standard est utilisée pour nos expériences¹.

¹A.K. Subramanian, Pytorch-VAE. <https://github.com/AntixK/PyTorch-VAE>, 2020.

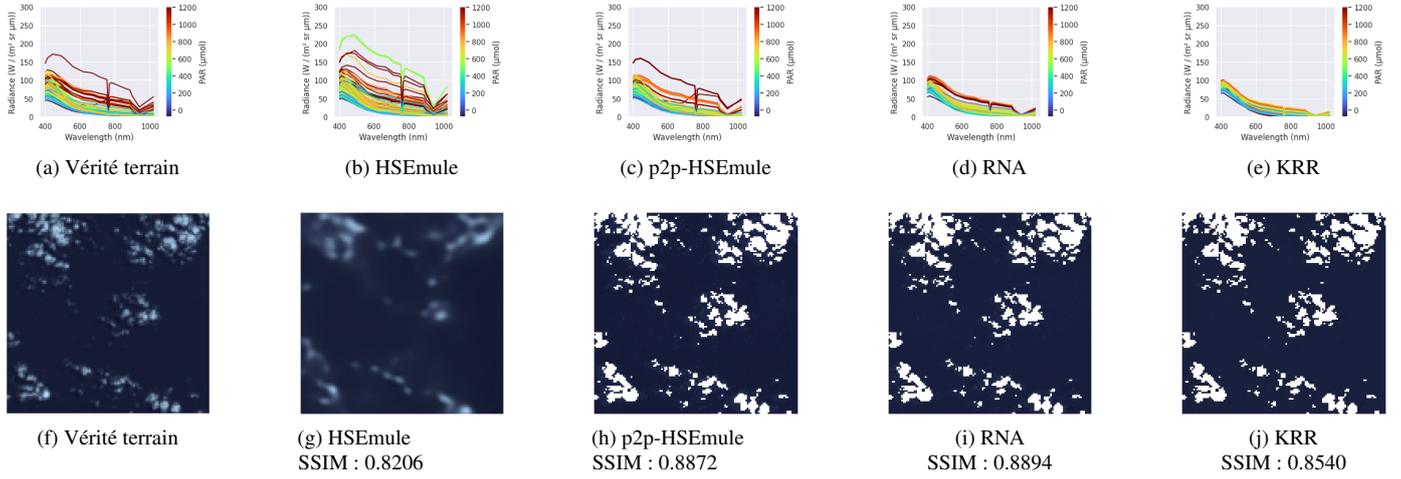


FIGURE 2 – En haut : Spectres Sentinel-3 originaux (a) et émulsés à l’aide de HSEmule (b), p2p-HSEmule (h), RNA (d) et KRR (e). En bas : Compositions RVB d’une image Sentinel-3 originale (f) et émulée avec HSEmule (g), p2p-HSEmule (h), RNA (i) et KRR (j). Les pixels de nuages sont masqués pour (h), (i) et (j) (en blanc).

3 Expériences et résultats

3.1 Jeu de données

Le jeu de données utilisé se compose d’images acquises par l’instrument Ocean and Land Color Instrument (OLCI) du satellite Sentinel-3 et des paramètres biophysiques associés. Les images sont acquises entre le 19 avril 2022 et le 23 avril 2022 au-dessus de l’océan Atlantique, de la mer Méditerranée, de la mer du Nord et de la Manche. Les images ont été recadrées à 128×128 pixels, tout en garantissant un minimum de 40% de pixels d’eau. 100 000 paires de paramètres biophysiques et d’images HS ont ainsi été obtenues. 25 000 et 5000 cubes ont respectivement été sélectionnés au hasard pour l’entraînement et pour la validation du modèle de VAE de la méthode HSEmule. L’entraînement du VAE du modèle p2p-HSEmule a été effectué avec 800 000 et 160 000 spectres pour son entraînement et sa validation, respectivement. Les paramètres biophysiques ont été choisis comme six variables indépendantes, à savoir, la concentration de chlorophyll-a (Chl), la concentration totale des matières en suspension (TSM), le coefficient d’atténuation de diffusion (K_d), le rayonnement photosynthétiquement actif (PAR), la colonne intégrée de vapeur d’eau (IWV) et le coefficient d’Angstrom (Å).

3.2 Résultats

Nous comparons les deux versions de notre approche proposée, baptisées HSEmule et p2p-HSEmule, aux méthodes de régression crête à noyau (KRR) et d’RNA qui imitent les spectres individuels. Ces méthodes sont issues des travaux de Verrelst *et al.* [10]. Les performances des émulateurs ont été évaluées sur un sous-ensemble de données de validation composé de 250 HSI, correspondant à 4 096 000 spectres. Les résultats quantitatifs sont présentés dans le tableau 1. Les métriques considérées sont les suivantes : l’erreur quadratique moyenne (EQM), la mesure de qualité de reconstruction spatiale SSIM (indice de similarité structurale), l’angle spectral (SA) et le coefficient de détermination R^2 . Notons que pour les trois méthodes d’émulation de spectres (p2p-HSEmule, RNA, KRR),

les pixels correspondant aux nuages ne sont pas émulsés et sont masqués lors des phases d’entraînement. Le calcul des métriques du tableau 1 ne les prend donc pas en compte.

TABLE 1 – Résultats obtenus et durées de traitement (en s).

	EQM	SSIM (moy)	SA (moy)	R^2	CPU (s)	GPU (s)
HSEmule	48.7950	0.8944	0.0491	0.8751	14.2600	0.5109
pix2pix-HSEmule	26.1081	0.8908	0.0372	0.7723	15.3800	0.6669
RNA	50.0067	0.8464	0.0588	0.6796	3.3271	0.6966
KRR	100.2100	0.7758	0.1177	0.4756	3.1305	-

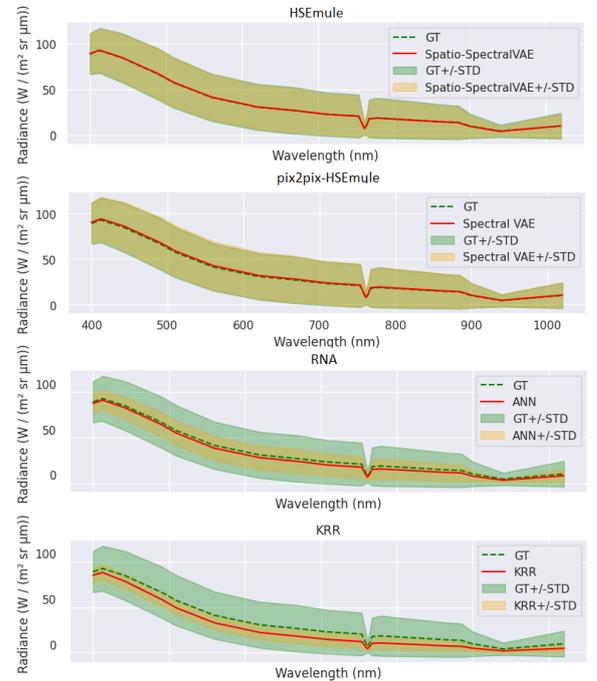


FIGURE 3 – Statistiques générales (moyenne, écart-type) pour l’évaluation et les spectres émulsés correspondants à l’aide de (de haut en bas) HSEmule, pix2pix-HSEmule, RNA et KRR.

Les résultats montrent que les deux versions proposées de notre approche surpassent à la fois les méthodes KRR et RNA

pour les métriques EQM, SSIM et SA, témoignant de la qualité de la reconstruction globale, spatiale et spectrale des cubes émulsés, respectivement. L'approche spectre-à-spectre, p2p-HSEmule, atteint les meilleurs résultats en terme d'EQM et de SA. L'approche de génération de cubes, HSEmule, quant à elle atteint le meilleur score de reconstruction spatiale (SSIM), mais aussi le meilleur score R^2 . Bien que les approches que nous proposons aient des temps de traitement CPU plus longs que KRR et RNA, ce à quoi nous nous attendions, elles sont nettement plus rapides lorsque les calculs sont effectués sur des GPUs, installés désormais de manière presque universelle dans les serveurs de calcul.

Les inspections visuelles, montrées dans la figure 2, confirmant que les scènes émulsées par HSEmule correspondent plus précisément à la couleur et aux caractéristiques spectrales des données Sentinel-3 originales, bien qu'un léger flou spatial soit noté. De plus, nous comparons les spectres émulsés par rapport à la vérité terrain dans la figure 2, (a) à (e). La couleur des spectres est affichée en fonction de la valeur du paramètre biophysique (PAR). Remarquons que les deux versions de notre méthode réussissent à capturer la relation entre les amplitudes des spectres et le PAR, même pour des valeurs de ce paramètre peu représenté dans le jeu de données d'entraînement. Les méthodes RNA et KRR n'arrivent pas à modéliser correctement cette relation. Globalement, HSEmule démontre une précision et une efficacité supérieures dans l'émulation d'images HS, en particulier dans les configurations prises en charge par le GPU, d'où l'intérêt d'utiliser une approche spatio-spectrale pour l'émulation. Bien que nous remarquons l'apparition de plus d'anomalies dans les spectres générées par cette approches.

Pour étudier plus en détail les performances de nos émulateurs, les statistiques générales (moyenne, écart-type) des spectres émulsés sur ceux des spectres de la vérité terrain sont affichées dans la figure 3. Cette figure montre une superposition presque parfaite entre les statistiques de nos méthodes et celles de la vérité terrain, en particulier pour l'approche spatio-spectrale HSEmule. L'approche spectre à spectre montre une faible surestimation en moyenne sur les bandes du visible (entre 400 et 750 nm). La méthode de la littérature RNA présente une faible sous-estimation en moyenne, sans pour autant capturer la variabilité des spectres du jeu de données. Enfin, l'approche KRR fournit les performances les moins bonnes. Ces résultats sont en concordance avec les scores R^2 du tableau 1.

4 Conclusion

Dans cette étude, nous avons comparé deux implémentations de notre approche couplée VAE-Interpolateur pour l'émulation d'images hyperspectrales, démontrant des améliorations substantielles en termes de précision, d'efficacité et de synthèse d'informations au niveau de la scène par rapport aux méthodes existantes, approche spatio-spectrale étant légèrement efficace à capturer la complexité d'une image hyperspectrale que la méthode spectrale. Les perspectives de ce travail incluent l'évaluation notre approche dans une application pratique telle que la récupération de paramètres biophysiques pour attester de la qualité d'exploitation scientifique des données émulsées. Plusieurs stratégies seront également étudiées pour atténuer le flou dû à l'utilisation d'une architecture VAE standard. Par

ailleurs, la validation de notre méthode sur des données plus complexes issues d'un simulateur de la végétation est envisagée.

5 Remerciements

Ces travaux sont financés par le pôle MTE de l'ULCO, les EUR MAIA (ANR-22-EXES-0009) et IFSEA (ANR-21-EXES-0011), et la région Hauts-de-France.

Références

- [1] A. ALBOODY, M. PUIGT, G. ROUSSEL, V. VANTREPOTTE, C. JAMET et T.-K. TRAN : DeepSen3 : Deep multi-scale learning model for spatial-spectral fusion of Sentinel-2 and Sentinel-3 remote sensing images. *In Proc. IEEE WHISPERS'22*, pages 1–5. IEEE, 2022.
- [2] J. M. AMIGO : *Hyperspectral imaging*. Elsevier, 2019.
- [3] C. BEN AZIZI, C. GUILLOTEAU, G. ROUSSEL et M. PUIGT : Coupled VAE and interpolator approach for fast hyperspectral image emulation. *In Proc. IEEE WHISPERS'24*, pages 1–5. IEEE, 2024.
- [4] A. BÖRNER, L. WIEST, P. KELLER, R. REULKE, R. RICHTER, M. SCHAEPMAN et D. SCHLÄPFER : Sensor : a tool for the simulation of hyperspectral remote sensing systems. *Photogramm. Eng. Remote Sens.*, 55(5-6):299–312, 2001.
- [5] J. L. GÓMEZ-DANS, P. E. LEWIS et M. DISNEY : Efficient emulation of radiative transfer codes using gaussian processes and application to land surface parameter inferences. *Remote Sensing*, 8(2):119, 2016.
- [6] K. HE, X. ZHANG, S. REN et J. SUN : Deep residual learning for image recognition. *In Proc. IEEE CVPR*, pages 770–778, 2016.
- [7] D. P. KINGMA et M. WELLING : Auto-encoding variational Bayes. arXiv e-print arXiv :1312.6114, 2022.
- [8] T. MOHANDOSS, A. KULKARNI, D. NORTHRUP, E. MWEBAZE et H. ALEMOHAMMAD : Generating synthetic multispectral satellite imagery from sentinel-2. *arXiv preprint arXiv :2012.03108*, 2020.
- [9] A. ODENA, V. DUMOULIN et C. OLAH : Deconvolution and checkerboard artifacts. *Distill*, 2016.
- [10] J. VERRELST, J. P. RIVERA CAICEDO, J. MUÑOZ-MARÍ, G. CAMPS-VALLS et J. MORENO : Scope-based emulators for fast generation of synthetic canopy reflectance and sun-induced fluorescence spectra. *Remote Sensing*, 9(9):927, 2017.
- [11] A. ZARE et K. HO : Endmember variability in hyperspectral analysis : Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.