



AfIA

Association française
pour l'Intelligence Artificielle

CNIA

Conférence Nationale en Intelligence Artificielle

PFIA 2024



Génération d'images de la Méiofaune à l'aide de StyleGAN2 : Cas des Copepoda

A. Martinel¹, A. Benzinou¹, K. Nasreddine¹, V. Foulon¹, C. Borremans², D. Zeppilli²

¹ ENIB, UMR CNRS 6285 LabSTICC, 29238 Brest, France

² Ifremer, 29280 Plouzané, France

Résumé

Nous explorons différentes approches hiérarchiques de transfert d'apprentissage d'un réseau antagoniste génératif StyleGAN2 afin de synthétiser des images de Copepoda. Il s'agit d'un des groupes les plus abondants de la faune aquatique, possédant peu d'images disponibles publiquement. Ces animaux sont de formidables bio-indicateurs de la pollution ou des changements environnementaux d'un milieu. Deux schémas d'apprentissage sont proposés. Le premier consiste à pré-entraîner le réseau avec les données d'un autre spécimen de même rang taxonomique et de faire un transfert d'apprentissage sur les données de l'animal étudié. Le deuxième consiste à pré-entraîner le réseau pour capturer les caractéristiques communes aux spécimens d'un rang taxonomique supérieur, pour enfin affiner le modèle au rang taxonomique inférieur souhaité. Ces méthodes visent à profiter des relations qui lient différents rangs taxonomiques. Les modèles obtenus sont ensuite évalués à l'aide des métriques FID et KID. Les images générées sont prometteuses, montrant des caractéristiques morphologiques typiques des copépodes. Ces données pourront ensuite être utilisées pour la formation de futurs taxonomistes et pour le développement de classifieurs de ces animaux, modèles qui nécessitent un grand nombre d'images pour leur entraînement.

Abstract

We explore two StyleGAN2 hierarchical transfer learning approaches in order to generate synthetic images of Copepoda animals. This is one of the most abundant groups in the aquatic fauna, yet only a few publicly available images are available. These animals are formidable bio-indicators of environmental changes or pollution of an habitat. Two learning approaches are proposed. The first is to pre-train the network with a dataset of another specimen of the same taxonomic rank and then fine tune it with the dataset of the studied animal. The second is to pre-train the network to capture features common to specimens of a higher taxonomic rank, finally refining the model to the desired lower taxonomic rank. Both methods aim to take advantage of the relationships that link different taxonomic ranks. The resulting models are then evaluated using the FID and KID metrics. The generated images are promising, showing typical morphological features of Copepods. These data can then be used to train future taxonomists and develop clas-

sifiers, models that require a large number of images for training.

Keywords

Apprentissage automatique, Réseaux adversariels génératifs, StyleGAN2, Meiofauna

1 Introduction

La méiofaune désigne l'ensemble des espèces animales d'une taille comprise entre $20\mu\text{m}$ et 1mm , en particulier les espèces vivant dans les sédiments au fond de l'eau sont dites benthiques. Les mots meiobenthos et méiofaune sont toutefois utilisés de façon ambiguë. Ces espèces jouent un rôle primordial au sein de l'écosystème marin profond, à la fois en tant que producteurs et consommateurs [14]. Elles réagissent plus rapidement que la macrofaune aux différentes formes de stress auxquelles elles sont exposées, faisant d'elles d'excellents bio-indicateurs des changements écologiques, de la pollution du milieu ou encore du dérèglement climatique : chalutage, décharge industrielle et agricole, pollution plastique, pollution aux métaux lourds et hydrocarbures, rejet de déchets nucléaires, déchets de munitions, exploitation minérale marine, acidification et désoxygénation des océans, etc. Avec 11 290 espèces identifiées [4], le groupe des Copepoda est le deuxième groupe le plus représenté de la méiofaune et également l'un des groupes les plus importants parmi les organismes planctoniques de cette taille. Les copépodes sont des animaux benthiques et pélagiques (vivant dans les sédiments et la colonne d'eau) dont la taille varie entre $250\mu\text{m}$ et 3cm . Comme pour d'autres groupes de la méiofaune, le nombre d'espèces inconnues est très important et les estimations varient énormément : entre 30 125 et 450 000 espèces de copépodes encore inconnues selon des études [12, 1]. La figure 1 montre quelques exemples d'images de copépodes.

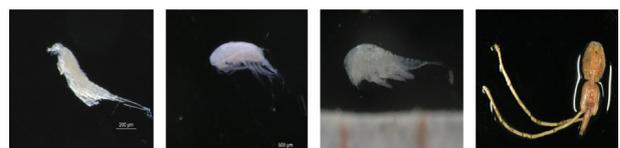


Figure 1: Exemples d'images de copépodes. De gauche à droite, un animal de l'ordre des Calanoida, Cyclopoida, Harpacticoida et Siphonostomatoida.

Néanmoins, l'identification taxonomique de la méiofaune demeure une tâche difficile. Elle repose sur l'utilisation de clefs taxonomiques et nécessite l'expertise de taxonomistes spécialistes des espèces étudiées, requérant plusieurs semaines de travail pour un échantillon de quelques centaines d'individus. Ce travail est rendu plus ardu par le manque de main d'œuvre, faisant du développement d'outils d'intelligence artificielle un enjeu majeur pour l'étude de la méiofaune. Cependant, ce développement est entravé par le manque de données. Motivés par les récentes améliorations apportées aux réseaux antagonistes génératifs (GANs) de haute définition entraînés sur de petits ensembles de données, nous étudions dans ce travail l'utilisation de réseaux neuronaux artificiels pour la génération de données synthétiques de copépodes.

Introduits pour la première fois en 2014, les GANs [5] sont devenus très populaires dans divers domaines de la génération de données, tels que la génération d'images, la traduction d'image à image ou de texte à image, l'augmentation des données ou encore le débruitage. Les GANs se composent de deux réseaux de neurones artificiels concurrents. Le premier, appelé générateur, crée à partir d'un bruit aléatoire de fausses données qui doivent tromper le second, appelé discriminateur. Ce dernier est entraîné à classifier entre données réelles et données générées. À la suite de la première architecture proposée [5], de nombreux nouveaux modèles ont vu le jour améliorant la qualité et la résolution des images générées. Récemment, deux GANs ont considérablement amélioré la qualité des images synthétiques ; les réseaux BigGANs [3] et StyleGANs [8] sont capables de générer des images de haute résolution, typiquement 512x512 pixels voir même 1024x1024 pixels. Le principal inconvénient des BigGANs est la taille importante de leur réseau, qui nécessite un très grand nombre d'images pour l'entraînement. Les améliorations récentes sur la seconde version du StyleGAN, à savoir le StyleGAN2 [9], pour les jeux de données limités nous ont conduit à choisir ce réseau pour base de notre travail. Nous proposons ici deux approches hiérarchiques de transfert d'apprentissage pour son entraînement à générer des images de Copépodes.

Le modèle, ses améliorations et les méthodes d'apprentissage que nous proposons sont expliqués plus en détail dans la section suivante. En sections 3 et 4, nous décrivons les données d'entraînement et les critères d'évaluation. Enfin, nous présentons et discutons les résultats obtenus dans la section 5.

2 Méthode proposée

Introduit en 2014, le DCGAN [10] est l'une des architectures convolutionnelles les plus populaires ; ce réseau est moins coûteux en terme de calcul et moins complexe à entraîner que d'autres GANs qui lui ont succédé. La particularité du StyleGAN est l'introduction de vecteurs de style : au lieu d'alimenter directement le générateur avec un vecteur de bruit gaussien, un réseau dense appelé réseau de mapping le transforme au préalable en un vecteur d'espace latent intermédiaire. Le vecteur latent obtenu est transformé

à son tour via des transformations affines en un vecteur de style **A**. Ce dernier est ensuite introduit dans chaque bloc du générateur. Une seconde version, StyleGAN2, améliore son prédécesseur, en augmentant ses performances et en évitant l'apparition d'artefacts. Les blocs de base du générateur de StyleGAN2 sont illustrés dans la Figure 2, le vecteur de style **A** est issu du réseau de mapping, dans notre cas un Perceptron de deux couches. Les variations stochastiques de l'image sont fournies par l'injection parallèle d'un vecteur de bruit **B**. La modulation met à l'échelle les poids des caractéristiques d'entrée en fonction du style **A** : $w_{ijkl} = s_i \cdot w_{ijkl}$, où w et w' sont respectivement les poids originaux et modulés, k et l sont les indices spatiaux, j est l'indice de la carte de caractéristique de sortie, et s_i est le style correspondant à la carte de caractéristiques d'entrée de rang i . Après cette opération, les activations de sortie ont un écart type de :

$$\sigma_j = \sqrt{\sum_{i,k,l} w'_{ijkl}{}^2} \quad (1)$$

Une opération de démodulation est appliquée :

$$w''_{ijkl} = \frac{w'_{ijkl}}{\sqrt{\sum_{i,k,l} w'_{ijkl}{}^2 + \epsilon}} \quad (2)$$

où ϵ est une constante qui permet d'éviter des instabilités numériques.

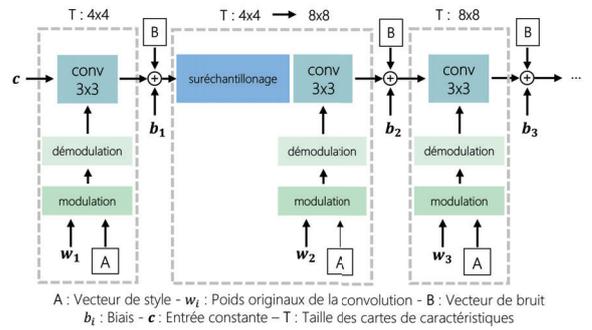


Figure 2: Schéma des trois premiers blocs du générateur du StyleGAN2. Les blocs suivants se succèdent jusqu'à atteindre la taille de l'image de sortie.

Enfin, nous proposons d'utiliser la limite adaptative d'augmentation de données ADA (Adaptive Discriminator Augmentation) [7] et la régularisation LeCam R_{LC} [13]:

$$R_{LC} = \mathbb{E}_{x \sim \tau} [\|D(x) - \alpha_F\|^2] + \mathbb{E}_{z \sim p_z} [\|D(G(z)) - \alpha_R\|^2] \quad (3)$$

où τ est le jeu de données d'apprentissage, p_z est la distribution préalable, G et D sont respectivement le générateur et le discriminateur, et enfin α_F et α_R sont deux moyennes mobiles exponentielles. Ces méthodes permettent de stabiliser l'apprentissage dans le cas de jeux de données limités.

Pour l'entraînement, nous proposons deux protocoles afin d'obtenir des modèles capables de génération d'images synthétiques selon le rang taxonomique souhaité. Notre première approche vise à bénéficier des connaissances apprises

sur un autre spécimen que celui étudié. Elle consiste à entraîner d'abord le réseau avec un jeu de données d'un autre spécimen de même rang taxonomique pour lequel nous avons plus d'images et ensuite avec le jeu de données de l'animal étudié. Cela permet d'apprendre les caractéristiques communes d'un spécimen de même rang taxonomique avant d'utiliser les images de l'animal à synthétiser pour spécifier ses propres caractéristiques. Par la suite, nous notons cette méthode "méthode latérale". Nous proposons une seconde approche, permettant au réseau de capturer les caractéristiques communes aux spécimens d'un rang taxonomique supérieur, pour enfin affiner le modèle au rang taxonomique inférieur souhaité. Cette méthode hiérarchique est appelée "méthode descendante", ainsi par exemple nous entraînons nos modèles avec l'ensemble des images de Copepodes avant d'affiner en n'utilisant seulement les images de Calanoides.

3 Données d'entraînement

Les données utilisées dans cette étude proviennent du jeu de données BOLD (The Barcode of Life Data System) [11]. Les images sont triées pour ne garder que les exemples de taille supérieure à 256 x 256 pixels et représentant l'animal en entier. Les images sont ensuite redimensionnées à 256 par 256 pixels. Les copépodes sont identifiés selon leur ordre taxonomique. L'ensemble de données est composé de 1969 Copepoda, dont 902 Calanoida, 766 Cyclopoida, 150 Harpacticoida et 151 Siphonostomatoida.

4 Métriques de performance

Pour évaluer la qualité des images générées, nous utilisons les mesures FID (Fréchet Inception Distance) [6] et KID (Kernel Inception Distance) [2]. Le FID mesure la différence entre les distributions statistiques des caractéristiques, générées à l'aide d'un modèle Inception pré-entraîné, des images synthétiques et réelles. Dans le cas du FID-50k, les caractéristiques de l'ensemble de données réelles sont comparées à l'aide de la distance de Fréchet aux caractéristiques de 50 000 images générées. Afin de valider nos méthodes d'entraînement, nous ajoutons une deuxième mesure à nos tests de méthodes d'apprentissage. En effet, Keras et al. [7] ont montré que le FID n'était pas une métrique idéale dans le cas de jeux de données limités. Dans ce cas, le KID qui est une mesure non biaisée est plus approprié pour juger de la qualité des images générées. Le KID est calculé à l'aide de l'écart moyen maximal (MMD) des caractéristiques générées (de la même manière que le FID), des images synthétiques et réelles. Les images générées de meilleure qualité présentent des valeurs FID et KID plus faibles.

5 Résultats et discussion

Afin de valider les performances des méthodes proposées, nous avons également entraîné et évalué un modèle DCGAN avec le même jeu de données Calanoida. Nous présentons dans le Tableau 1 les résultats quantitatifs des modèles DCGAN et StyleGAN2. L'architecture DCGAN a du mal à

générer des images suffisamment réalistes pour s'apparenter à des images réelles. Le StyleGAN2 en revanche obtient de bien meilleurs résultats. Ayant validé l'avantage du StyleGAN2 par rapport au DCGAN, nous expérimentons la limite adaptative d'augmentation de données ADA et la régularisation LeCam qui améliorent toutes deux ses performances.

Table 1: Comparaison des performances des modèles entraînés avec le sous-ensemble des données Calanoida.

Model	Fid50k ↓
DCGAN [10]	239.240
StyleGAN2 [9]	137.426
StyleGAN2 + ADA [7]	44.774
StyleGAN2 + ADA + R_{LC} [13]	39.341



Figure 3: Images synthétiques de copépodes générées par la méthode hiérarchique descendante. 1^{ère} ligne : Calanoida, 2^{ème} : Cyclopoida, 3^{ème} : Harpacticoida et 4^{ème} : Siphonostomatoida.

Dans le tableau 2, nous présentons les mesures de qualité des trois méthodes. Sans surprise, plus nous avons de données, meilleure est la qualité des images générées ; les modèles entraînés sur le jeu de données Calanoida montrent ainsi les meilleurs résultats. Nous observons que pour chaque spécimen, l'apprentissage par transfert descendant depuis les copépodes permet au modèle de générer des données de meilleure qualité. Cette approche hiérarchique de l'apprentissage par transfert, en commençant par l'entraînement au rang taxonomique plus général (ici Copepoda), puis en affinant le modèle au rang taxonomique plus spécifique semble être la méthode la plus efficace. Notons les différences de perception de qualité d'image des deux métriques utilisées ; en effet, lorsque les modèles entraînés sur les Cyclopoida et les Harpacticoida ont respectivement des KID proches, leurs FID respectifs sont très éloignés. Cette différence est selon nous due aux différences de taille et de diversité des deux jeux de données. En effet le jeu de

Table 2: Comparaison des performances des méthodes d’entraînement utilisées (directe, latérale et descendante). Pour chaque ordre nous entraînons sur son jeu de données, un modèle sans transfert avec les poids initialisés de manière aléatoire, un modèle pré-entraîné sur les données de Calanoida et un modèle pré-entraîné sur les données regroupées de Copepoda.

Ordre	Méthode directe		Méthode latérale		Méthode descendante	
	Fid50k ↓	Kid50k ↓ ($\times 10^2$)	Fid50k ↓	Kid50k ↓ ($\times 10^2$)	Fid50k ↓	Kid50k ↓ ($\times 10^2$)
Calanoida	39.341	1.204	X	X	36.529	0.931
Cyclopoida	60.401	2.327	50.504	1.545	47.017	1.215
Harpacticoida	119.298	2.323	114.288	2.070	108.533	1.503
Siphonostomatoida	111.865	1.611	107.610	1.549	96.995	0.990

données des Harpacticoida est beaucoup plus petit (150 images) que celui des Cyclopoida (766 images) et contient des images moins diversifiées (posture, microscope utilisé, etc). Ces mesures de performance ne quantifient pas forcément l’intégralité des traits caractéristiques de chaque spécimen. Afin de compléter ces résultats et valider la capacité de nos images à rendre compte des différences entre les ordres, nous avons présenté nos images à des experts mondiaux des Copépodes. Ces derniers ont confirmé la qualité de nos images. Des exemples d’images générées par nos modèles (entraînés via la méthode descendante) sont présentés dans la Figure 3. Les images générées présentent des caractéristiques complexes comme les réglettes, des réflexions lumineuses ou même des œufs.

6 Conclusion et perspectives

Nous avons proposé deux approches d’apprentissage par transfert hiérarchique pour la génération d’images de copépodes à l’aide d’un réseau StyleGAN2. Nous obtenons des résultats très prometteurs, à l’aide de nos deux stratégies les modèles utilisés ont réussi à générer des images d’animaux qui ont été identifiés comme des copépodes par des taxonomistes. Ces approches peuvent être utilisées sur d’autres espèces de la méiofaune afin de générer des données synthétiques de spécimens pour lesquels peu de données sont disponibles. Ces données synthétiques pourraient constituer un nouvel outil pour former des taxonomistes ou être utilisées pour l’apprentissage de réseaux classifieurs. Une lecture par des systématistes taxonomistes a permis de valider la qualité des images synthétisées d’un point de vue taxonomique. En perspective, nous envisageons d’introduire de l’information de haut niveau dans les premières couches du réseau, comme les traits fonctionnels morphologiques systématiquement utilisés par les taxonomistes ainsi que de tester la troisième version du StyleGAN, censée améliorer la représentation interne des images.

References

- [1] W. Appeltans, S. T. Ahyong, G. Anderson, M. V. Angel, T. Artois, N. Bailly, et al. The magnitude of global marine species diversity. *Current Biology*, 22(23):2189–2202, 2012.
- [2] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton. Demystifying mmd gans. *ICLR*, 2018.
- [3] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096, 2018.
- [4] M. J. Costello and C. Chaudhary. Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Current Biology*, 27(11):R511–R527, 2017.
- [5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al. Generative adversarial networks. *NIPS*, page 2672–2680, 2014.
- [6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [7] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12104–12114. Curran Associates, Inc., 2020.
- [8] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018.
- [9] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. *CoRR*, abs/1912.04958, 2019.
- [10] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2014.
- [11] S. Ratnasingham and P. Hebert. Bold: The barcode of life data system (www.barcodinglife.org). *Molecular ecology notes*, 7:355–364, 06 2007.
- [12] S. Seifried. Phylogeny of harpacticoida (copepoda): Revision of ‘maxillipedasphalea’ and exanechentera. *Cuvillier Verlag, Göttingen Germany Ph.D. thesis from 2002*, pages :1–259., 2003. The World Of Copepods (T. Chad Walter).
- [13] H. Tseng, L. Jiang, C. Liu, M. Yang, and W. Yang. Regularizing generative adversarial networks under limited data. *CoRR*, abs/2104.03310, 2021.
- [14] D. Zeppilli, J. Sarrazin, D. Leduc, P. Martinez Arbizu, D. Fontaneto, C. Fontanier, et al. Is the meiofauna a good indicator for climate change and anthropogenic impacts? *Marine Biodiversity*, 45, 09 2015.