

3 Juillet 2024 - CNIA 2024, La Rochelle

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

MARTINEL Anthonin* , BENZINOUE Abdesslam* , NASREDDINE Kamal* ,
FOULON Valentin* , BORREMANS Catherine† and ZEPILLI Daniel†

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

† Ifremer, 29280 Plouzané, France

Email: martinel@enib.fr

1. Contexte et problématique

La Méiofaune

Des bio-indicateurs

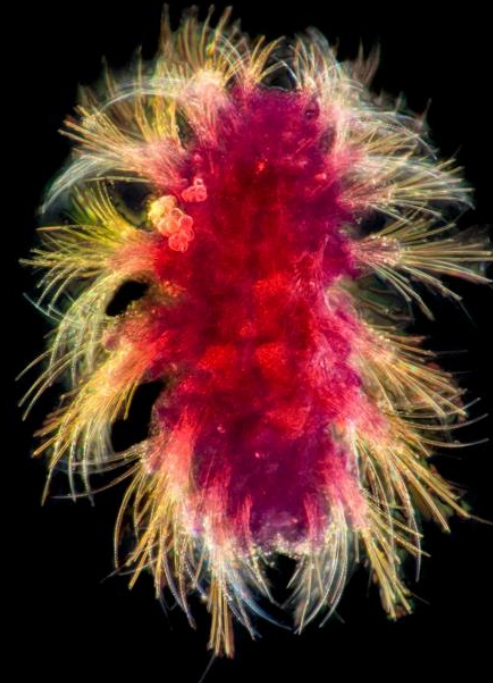
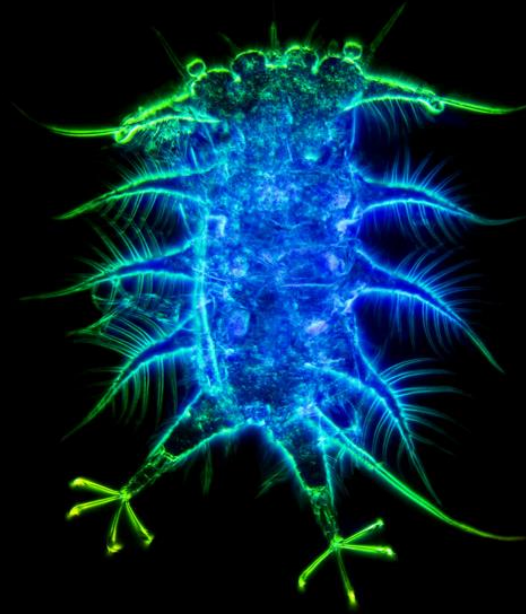
Meio (« *μειου* ») : plus petit

Echelle de taille : 20μm → 1mm

Copépodes :

Echelle de taille : 250μm → 3cm

Espèces inconnues : 30,125 → 450,000



1. Contexte et problématique

Générer des données synthétiques d'animaux à partir de petits jeux de données

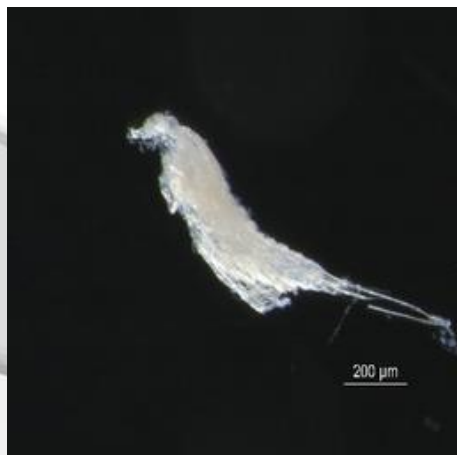
- Nouveaux outils pour les taxonomistes
- Jeux de données synthétiques pour former des classifieurs à l'avenir

2. Jeux de données

Calanoida



Cyclopoida



Harpacticoida



Siphonostomatoida

Nom	Nombre d'images
Calanoida	902
Cyclopoida	766
Harpacticoida	150
Siphonostomatoida	151
Total Copepoda	1969

Jeu de données BOLD

3. StyleGAN2 (+ ADA + LeCam)

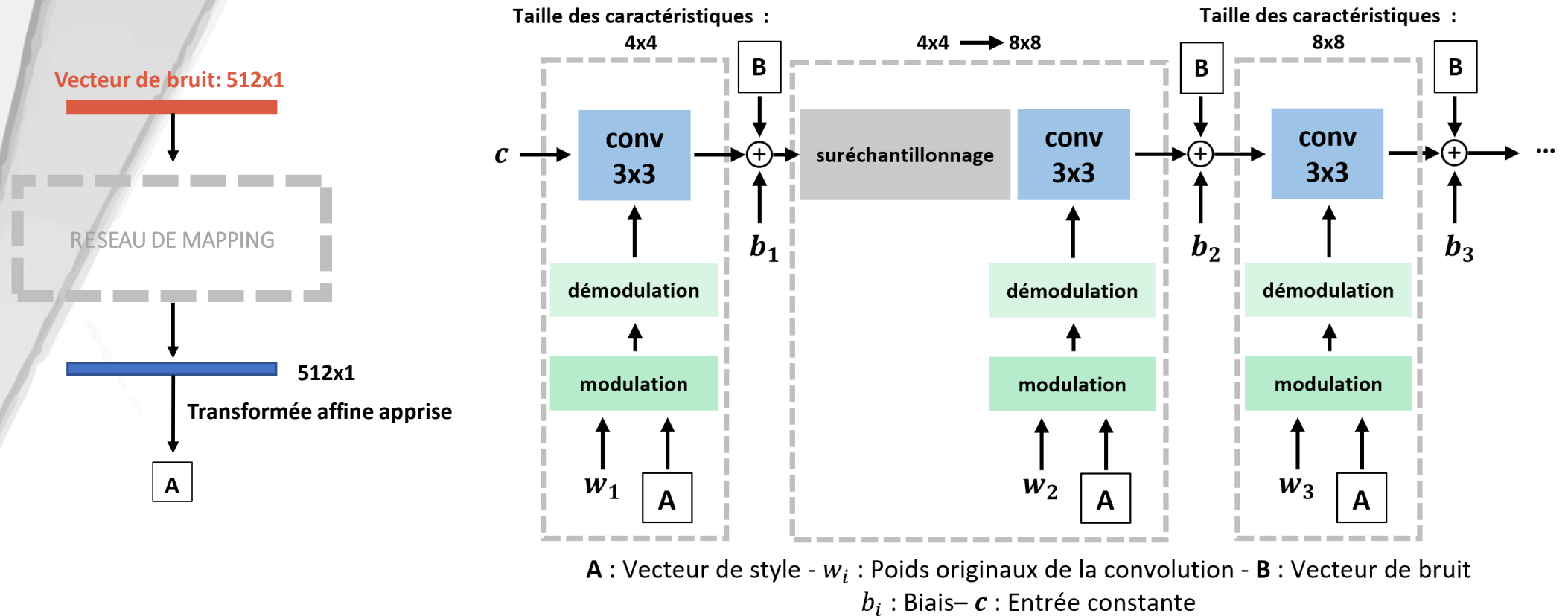
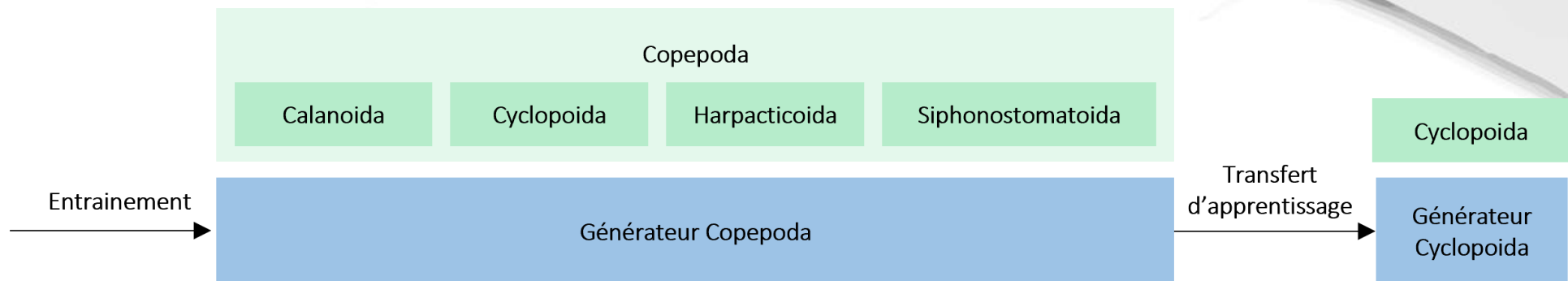
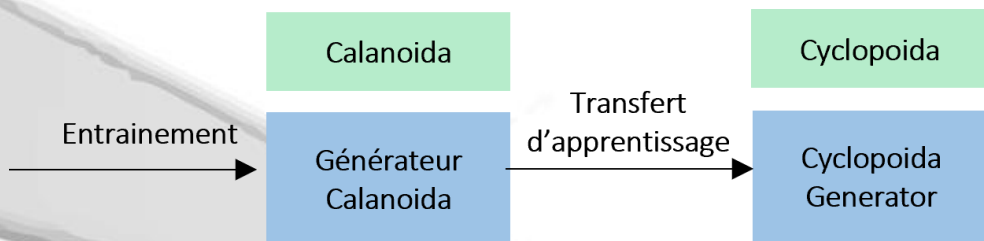


Schéma des trois premiers blocs du générateur du StyleGAN2. Les blocs suivants se succèdent jusqu'à atteindre la taille de l'image de sortie

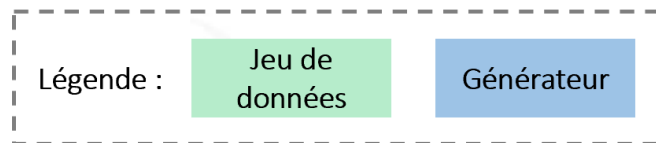
4. Deux approches par transfert d'apprentissage hiérarchiques



Transfert d'apprentissage depuis le réseau entraîné à générer des images de Copépodes (Approche descendante)



Transfert d'apprentissage depuis le réseau entraîné à générer des images de Calanoides (Approche latérale)



5. Métriques d'évaluation

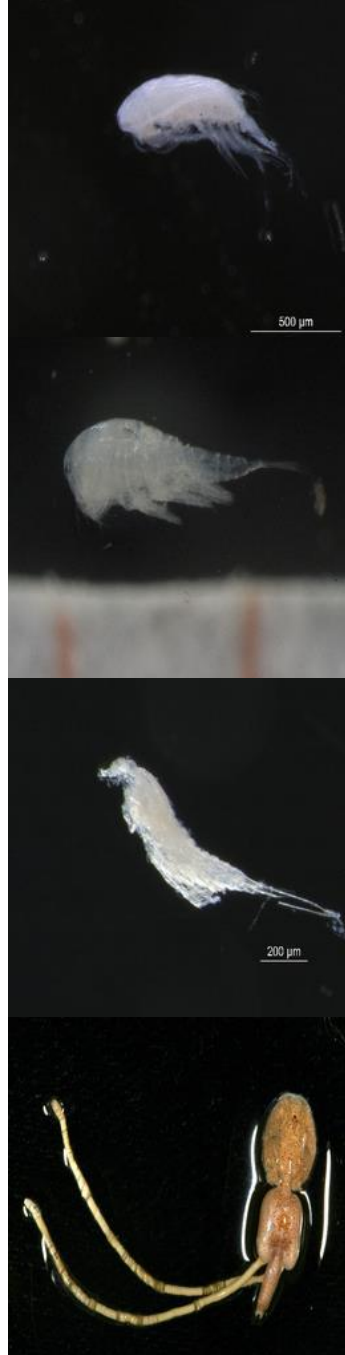
Fréchet Inception Distance (FID)
et Kernel Inception Distance (KID)

$$FID = \|\mu_r - \mu_g\|_2^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g})$$

- (μ_r, Σ_r) : Distribution gaussienne ajustée aux caractéristiques **réelles**
- (μ_g, Σ_g) : Distribution gaussienne ajustée aux caractéristiques **générées**

Le *KID* est calculé à l'aide de l'**écart moyen maximal** (MMD) des caractéristiques réelles et générées à l'aide d'un réseau Inception

6. Résultats et discussion



Images réelles

Calanoida

Cyclopoida

Harpacticoida

Siphonostomatoida



Images générées

6. Résultats et discussion

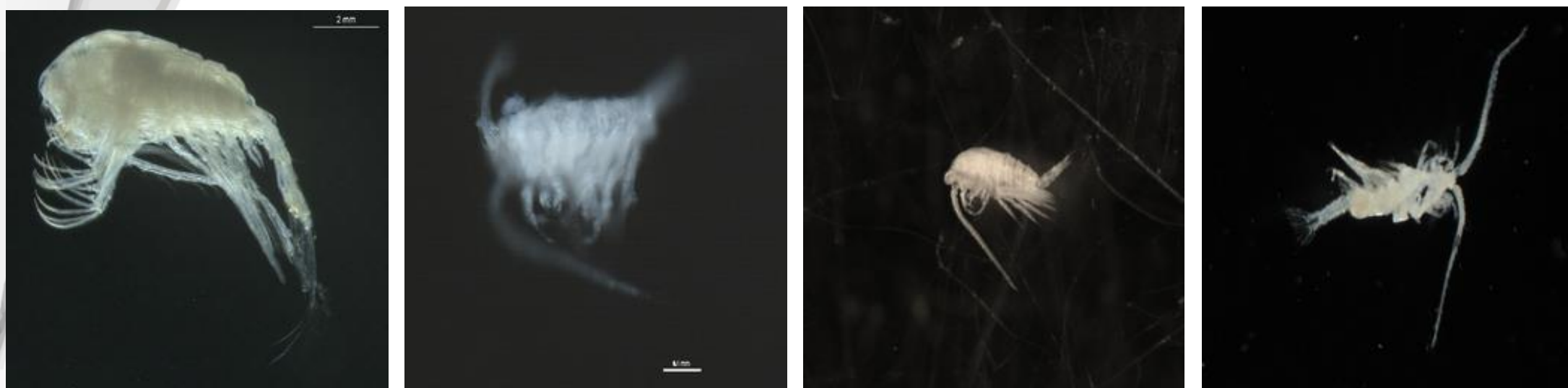
Ordre	Méthode directe		Méthode latérale		Méthode descendante	
	Fid50k ↓	Kid50k ↓ ($\times 10^2$)	Fid50k ↓	Kid50k ↓ ($\times 10^2$)	Fid50k ↓	Kid50k ↓ ($\times 10^2$)
Calanoida	39.341	1.204	X	X	36.529	0.931
Cyclopoida	60.401	2.327	50.504	1.545	47.017	1.215
Harpacticoida	119.298	2.323	114.288	2.070	108.533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

7. Conclusion

- Deux approches d'apprentissage par transfert hiérarchique :
 - Une approche latérale
 - Une approche descendante
- **Résultats validés par des spécialistes de la Méiofaune**
- Un nouvel outil pour former les futurs taxonomistes
- Perspectives :
 - Utiliser ces données pour entrainer des classifieurs
 - Essayer le StyleGAN3
 - Générer des images à des rangs taxonomiques plus bas



Exemples d'images synthétiques de Calanoida générées à l'aide de l'approche descendante



Exemples d'images de Calanoida provenant du jeu de données d'entraînement

Validation des résultats par des spécialistes

- Nous présentons à des taxonomistes spécialistes des Copépodes un mélange d'images réelles et générées
- A l'aide d'un indice de confiance, nous leur demandons de classer ces images :
 - **D'abord selon leur ordre taxonomique** (sans qu'ils soit mention que certaines images sont générées)
 - Puis **entre image générées et réelles**

On peut ainsi comparer les différences de résultats entre les images réelles et générées :

Les images synthétiques peuvent elles tromper les spécialistes ?

Les images synthétiques sont elles des chimères ?

Validation des résultats par des spécialistes

Real/Fake↓ /Order→	Calanoida	Cyclopoida	Harpacticoida	Siphonostomatoida
Réelle	85,1	52,8	100,0	58,3
Générée	79,4	40,7	53,3	22,2

Résultats de la première expérience : proportion (%) de bonne classification entre les différents ordres taxonomiques selon les images réelles et générées

Experts↓/GT→	Réelle	Générée	Inconnue
Réelle	87	49	0
Générée	43	29	0
Inconnue	1	1	0

Résultats de la deuxième expérience : matrice de confusion