

COMPTE RENDU

17/03/2024



L'ESTIMATION DE L'ÂGE DES OTOLITHES PAR DES APPROCHES D'APPRENTISSAGE AUTOMATIQUE

Groupe:

Achiba Ahmed
Bougra Mohamed
EL Quartl Mouad
El Okri Amine
Zahdi Ghita

encadré par:

Emilie Poisson-Caillault

2023/2024



github:

<https://github.com/amineelokri/Estimation-de-l-age-des-poissons-par-Deep-Learning.git> .

branche master

Division de la base donnée:

La base de données se divise en deux catégories : les images "left_and_right" et les images "right". Dans un premier temps, nous travaillerons sur les images des otolithes droits pour prédire l'âge des poissons. C'est pourquoi nous avons divisé les métadonnées en deux catégories.

Un problème rencontré lors de cette division des métadonnées est que certaines références sont présentes dans les métadonnées, mais les images correspondantes sont absentes. Ou inversement, certaines images sont présentes, mais les données associées manquent dans les métadonnées.

Dans la nouvelle métadonnée, nous avons conservé uniquement l'âge et la référence de l'image, puisque ce sont les seules colonnes dont nous aurons besoin pour entraîner notre modèle.

Visualisation de la base de donnée et statistique:

Dimension de la base de données:

La base de données contient 5879 lignes et 2 colonnes qui sont : l'âge et la référence de l'image .

```
#Les dimensions du base de donnees  
data.shape
```

```
(5879, 2)
```

```
#Les columns, on a besoin seulement de l'age e  
data.columns
```

```
Index(['Reference_PC', 'Age'], dtype='object')
```

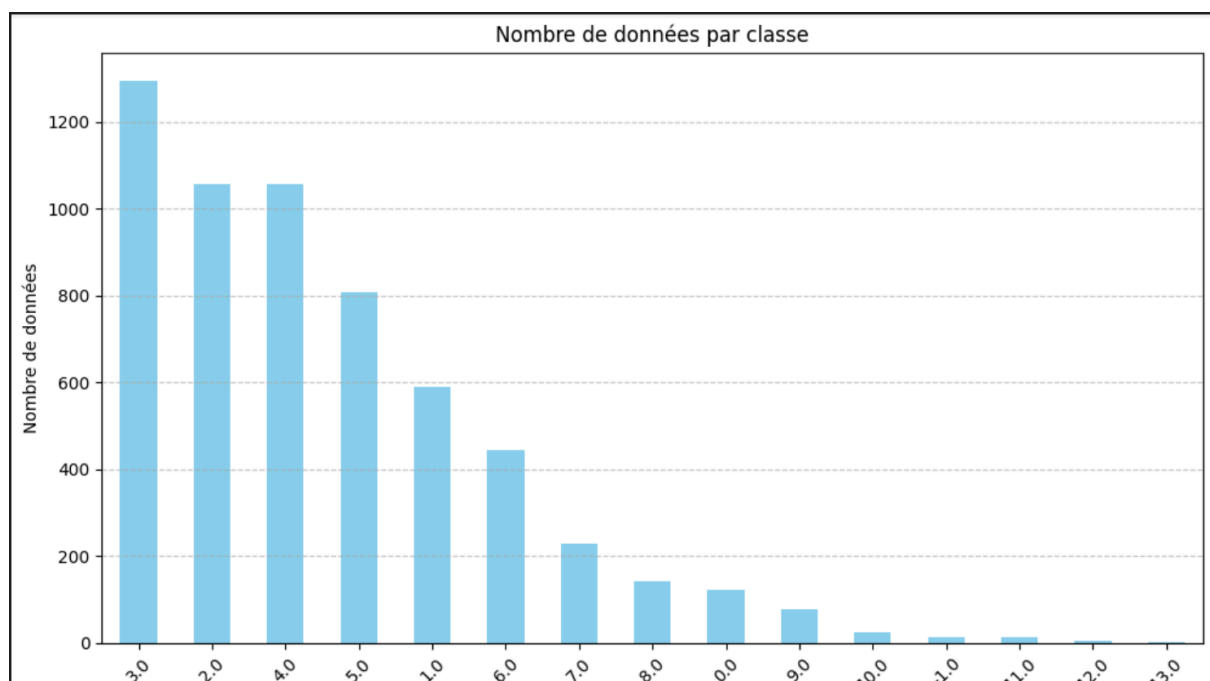
Les classes:

La base de données contient 15 classes, chaque classe représente l'âge d'une catégorie des otolithes et la classe -1 représente l'ensemble des otolithes ayant un âge inconnu.

```
#Les classes
classes = data['Age'].unique()
#on remarque que les classes varient de -1 à 13 ou -1 signifie que l'âge est inconnu
classes

array([ 5.,  4.,  3.,  8.,  7.,  6.,  2.,  9.,  1., -1.,  0., 10., 11.,
        12., 13.])
```

Nombre des otolithes par classe:



On observe un déséquilibre dans la distribution des données entre les classes, où la majorité des données ont un âge de 2, 3 et 4. Ceci peut potentiellement entraîner un surajustement (overfitting) du modèle.

Dimensions des images:

```
[(718, 534), (1161, 790), (1301, 822), (1315, 806), (1281, 815), (1207, 836), ...]
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Nous constatons que les images ont des dimensions variées. De plus, il est observé que les poissons plus âgés ont généralement des otolithes de plus grandes dimensions, ce qui est cohérent, puisque le nombre d'anneaux et l'âge sont proportionnels. Par conséquent, nous devons prendre en compte les différences de dimensions des images dans notre traitement.

Les couches de convolution des CNN effectuent des opérations qui dépendent de la taille de la fenêtre de convolution et de la taille du filtre. Si les images ont des dimensions différentes, ces opérations ne peuvent pas être appliquées de manière cohérente sur toutes les images, ce qui entraîne des erreurs ou des problèmes d'ajustement du modèle.

Cependant, nous devons redimensionner les images à une taille fixe avant de les utiliser pour l'entraînement du modèle. Cela garantit que toutes les images ont les mêmes dimensions et peuvent être traitées de manière cohérente par le CNN.

Statistique de la distribution des âges des otolithes dans la base de données:

count	5879.000000
mean	3.647049
std	1.997388
min	-1.000000
25%	2.000000
50%	3.000000
75%	5.000000
max	13.000000

count (nombre) : Il y a un total de 5879 entrées dans la base de données, ce qui représente le nombre d'otolithes dans l'échantillon.

mean (moyenne) : La moyenne des âges des otolithes est d'environ 3.65. Cela indique que l'âge moyen des otolithes dans l'échantillon est autour de 3.65 ans.

std (écart-type) : L'écart-type des âges des otolithes est d'environ 1.997. Cela indique que les valeurs sont dispersées autour de la moyenne.

25% (premier quartile) : Le premier quartile se situe à un âge de 2 ans. Cela signifie que 25% des otolithes ont un âge égal ou inférieur à 2 ans.

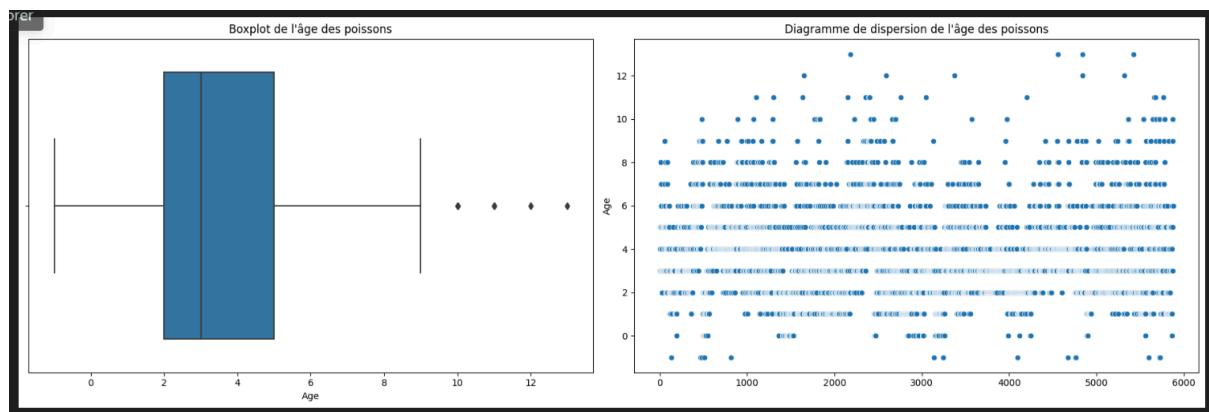
50% (médiane) : La médiane des âges des otolithes est de 3 ans. Cela signifie que la moitié des otolithes ont un âge inférieur ou égal à 3 ans, et l'autre moitié ont un âge supérieur à 3 ans.

75% (troisième quartile) : Le troisième quartile se situe à un âge de 5 ans. Cela signifie que 75% des otolithes ont un âge égal ou inférieur à 5 ans.

Nous faisons face à une base de données où les classes ne sont pas équilibrées.

max (maximum) :

L'âge maximum des otolithes est de 13 ans. Cela représente l'otolithe le plus ancien dans l'échantillon.



Data preprocessing :

Nous devons redimensionner les images à une taille fixe pour entraîner notre modèle. Cependant, nous avons rencontré un problème lors de la manipulation des images : celles-ci ont une extension '.tif', tandis que Keras ne prend en charge que les extensions ".jpeg", ".jpg", ".png", ".bmp" et ".gif". Bien que nous ayons trouvé une seule bibliothèque, "tiffle", qui permet de lire ce type d'images, nous avons rencontré des problèmes lors du prétraitement des images, tels que le redimensionnement et la manipulation des images. Pour surmonter cette difficulté, nous avons décidé de convertir les images des otolithes droits du format '.tif' au format '.png' en utilisant la bibliothèque 'imageio'. Cette conversion nous permet d'utiliser Keras et de simplifier l'entraînement des données.

```
dossier='./right/'
out_folder='./right_png/'
imgs=[]
index=""

for e in data['Reference_PC']:
    file=dossier+e+'.tif'
    out_file=out_folder+e+'.png'
    index=e
    img = tiff.imread(file).transpose(1,2,0)
    imageio.imwrite(out_file, img)
    #resized_img=cv2.resize(img,(256, 256),interpolation=cv2.INTER_AREA

    imgs.append(img)

img=None
```