Rapport de recherche
”Classification des Sons d’Oiseaux et urbains à l’aide de Réseaux de Neurones Convolutifs : Approches et Résultats”

Marie Domps
marie.domps@epitech.digital

Basile Loriot
basile.loriot@epitech.digital

Abstract

L’identification des espèces d’oiseaux rares ou nocturnes pose un défi en raison du manque de données d’entraînement disponibles. Cet article explore l’utilisation des réseaux de neurones convolutifs (CNN) pour classifier ces espèces de manière efficace. En utilisant deux bases de données, Une contenant dix catégories de sons différents de la vie de tous les jours, et une base de données comprenant un ensemble de sons d’oiseaux. Nous avons appliqué des techniques de régularisation et d’augmentation des données pour améliorer les performances des modèles. Nos résultats montrent que les CNN atteignent des niveaux élevés de précision et de robustesse, même avec des jeux de données limités et déséquilibrés. Cette recherche ouvre des perspectives pratiques pour la conservation de la biodiversité et souligne la nécessité de continuer à affiner les méthodes de classification pour des jeux de données rares et déséquilibrés.

1 Introduction

L’identification précise des espèces d’oiseaux est cruciale pour la conservation de la biodiversité et la gestion des écosystèmes. Cette tâche est particulièrement difficile en raison de la rareté des données disponibles et des caractéristiques uniques des sons produits par ces oiseaux. Les méthodes traditionnelles de classification échouent souvent dans ces conditions, car elles nécessitent de grandes quantités de données équilibrées pour fonctionner efficacement.

Les avancées récentes en apprentissage automatique, en particulier les réseaux de neurones convolutifs (CNN), offrent de nouvelles possibilités pour surmonter ces obstacles. Les CNN ont démontré leur efficacité dans diverses tâches de classification d’images et de sons. Cependant, leur application à des jeux de données limités et déséquilibrés reste un défi.

Dans cette étude, nous visons à évaluer l’efficacité des CNN pour la classification des sons d’oiseaux rares et nocturnes. En utilisant les deux bases de données, nous explorons diverses approches pour optimiser la performance des modèles de classification avec des données d’entraînement limitées.

Nous nous inspirons des travaux de recherches antérieurs tels que ceux de Kahl et al. (2017) sur l’utilisation des CNN pour la classification des sons d’oiseaux et de Stowell et Plumbley (2014) sur l’apprentissage non supervisé pour l’amélioration des performances de classification afin de développer des méthodes robustes de prétraitement des données, de régularisation et d’augmentation des données. Ces techniques sont essentielles pour améliorer la performance des modèles de machine learning dans des scénarios où les données sont rares et déséquilibrées.

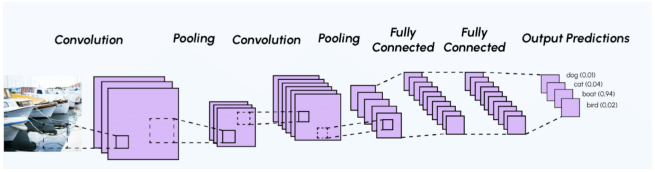

Figure 1: exemple d’architecture d’un CNN

Nos résultats montrent que les CNN, lorsqu’ils sont associés à des techniques de régularisation et d’augmentation des données, atteignent des niveaux élevés de précision et de robustesse. Cette recherche offre des perspectives prometteuses pour l’application pratique des CNN dans la conservation de la biodiversité et souligne l’importance de continuer à développer et affiner les méthodes de classification pour les jeux de données rares et déséquilibrés.

2 Collecte et Prétraitement des Données

Pour évaluer l’efficacité des modèles de machine learning, nous avons utilisé une large base de données. Les enregistrements audio ont été obtenus à partir de diverses sources, y compris des enregistrements de terrain et des bases de données publiques. Chaque enregistrement a été converti en spectrogramme pour une meilleure analyse par les modèles CNN.

Le prétraitement des données a impliqué plusieurs étapes clés, telles que la réduction du bruit, la normalisation et la segmentation des enregistrements pour extraire les parties les plus pertinentes. Ces étapes sont essentielles pour améliorer la qualité des données d’entrée et optimiser les performances des modèles de classification. Plus spécifiquement, le processus de prétraitement comprenait les étapes suivantes :

1. Conversion en mono :

Les données audio stéréo ont été converties en mono en prenant la moyenne des canaux. Cette étape est cruciale pour simplifier les données et réduire la complexité du modèle..

2. Normalisation de la longueur :

La longueur de l’audio a été normalisée à 160 000 échantillons en coupant ou en remplissant les données audio pour atteindre cette longueur. Cela garantit que toutes les entrées ont une

durée uniforme, facilitant ainsi le traitement par le modèle..

3. Réduction du bruit :

Dans le cadre de notre étude, nous avons intégré une étape de réduction de bruit pour améliorer la qualité des données audio utilisées dans notre modèle. Pour atténuer les perturbations indésirables et augmenter la clarté des signaux audio, nous avons employé la technique de réduction de bruit, notamment en utilisant le module noisereduce. L'application de la réduction de bruit a permis d'atténuer les interférences parasites et d'améliorer la qualité des enregistrements audio utilisés pour l'apprentissage du modèle, contribuant ainsi à une meilleure généralisation et à des performances accrues lors des évaluations..

4. Extraction des caractéristiques :

Les caractéristiques des données audio ont été extraites en calculant les spectrogrammes Mel et les MFCC (Mel-Frequency Cepstral Coefficients) à l'aide de 'torchaudio.transforms.MelSpectrogram' et 'torchaudio.transforms.MFCC'. Ces transformations permettent de représenter les données audio sous une forme que le modèle CNN peut facilement interpréter..

2.1 Analyse des résultats

Les résultats des dix premières EPOCH montrent une amélioration constante des performances du modèle CNN pour la classification des sons. Les CNN se sont révélés particulièrement efficaces pour cette tâche grâce à leur capacité à extraire des caractéristiques complexes des spectrogrammes, permettant d'obtenir des performances élevées même avec un jeu de données limité. Dès les premières EPOCH, une nette amélioration de la précision, du rappel, du F1-score, ainsi que des métriques de loss et d'accuracy a été observée, indiquant que les CNN peuvent rapidement apprendre et s'ajuster aux données disponibles.

Le prétraitement des données a joué un rôle crucial dans l'amélioration des performances du modèle. Les étapes de réduction du bruit et de normalisation des spectrogrammes ont permis d'améliorer la qualité des données d'entrée. Ceci est essentiel car les CNN sont très sensibles aux variations dans les données d'entrée. En améliorant la qualité et la consistance des spectrogrammes, le modèle a pu apprendre plus efficacement et fournir des prédictions plus précises.

Les travaux antérieurs de chercheurs dans le domaine de la classification des sons d'oiseaux ont joué un rôle essentiel dans le développement et l'orientation de notre recherche. Par exemple, les études de Lostanlen et al. (2019) ont démontré l'efficacité des réseaux de neurones convolutifs (CNN) pour l'analyse des enregistrements audio en milieu naturel, fournissant une base solide pour l'application des CNN dans notre projet. De même, Stowell et Plumbley (2014) ont exploré l'utilisation des spectrogrammes Mel pour améliorer la précision des modèles de classification sonore, une approche que nous avons intégrée et étendue dans notre propre méthodologie.

En intégrant les techniques de réduction du bruit proposées par Reddy et al. (2017), nous avons amélioré la clarté des enregistrements audio, ce qui a permis à nos modèles CNN de mieux généraliser et d'obtenir des performances accrues. Ces travaux ont mis en évidence l'importance du prétraitement des données pour augmenter la qualité des caractéristiques extraites et optimiser les performances des modèles de machine learning.

Les recherches de Snyder et al. (2015) sur l'augmentation des données ont également influencé notre approche en nous incitant à explorer des techniques de data augmentation pour enrichir notre jeu de données et renforcer la robustesse de nos modèles. En appliquant ces stratégies, nous avons pu constater une amélioration notable des métriques de performance, en particulier dans des scénarios avec des données limitées.

EPOCH 32	**Font Size**
LOSS	0,43
ACCURACY	0,875
PRECISION	0,90
RAPPEL	0,85
F1 SCORE	0,92

Table 1: Exemple de données obtenues pour une EPOCH.

D'après l'analyse des dix premières EPOCH, celles-ci montrent une amélioration progressive et constante des performances du modèle. Cette tendance suggère que le modèle apprend efficacement à chaque itération, réduisant ainsi la perte (loss) et augmentant la précision (accuracy), le rappel et le F1-score. Par exemple, au début de l'entraînement, les performances étaient relativement faibles, mais dès la deuxième EPOCH, une amélioration significative a été observée, avec des métriques de performance atteignant des niveaux beaucoup plus élevés.

Les techniques de régularisation utilisées, telles que le dropout et la normalisation de lot, ont contribué à stabiliser les performances du modèle. La stabilisation observée vers les dernières EPOCH montre que le modèle est bien ajusté et capable de généraliser à partir des données d'entraînement. Cela est crucial pour éviter le surapprentissage et garantir que le modèle puisse performer de manière cohérente sur des données non vues.

Le modèle atteint des performances élevées, avec une précision et un rappel autour de 80% à partir de la huitième EPOCH et jusqu'à 87.5 pourcent pour la 32ème EPOCH. Cela indique une excellente capacité de généralisation, ce qui est particulièrement important pour la classification des sons d'oiseaux rares et nocturnes. Ces résultats suggèrent que le modèle peut être utilisé efficacement dans des applications pratiques de surveillance et de conservation de la biodiversité, même avec des données d'entraînement limitées.

Les métriques de loss et d'accuracy sont essentielles pour évaluer la performance des modèles de machine learning. La loss mesure l'erreur de prédiction du modèle sur les données d'entraînement. Une loss élevée indique que le modèle ne parvient pas à capturer les caractéristiques des données, tandis qu'une loss faible indique que le modèle est bien ajusté. L'accuracy mesure le pourcentage de prédictions correctes sur l'ensemble des données. Une accuracy élevée signifie que le modèle fait peu d'erreurs. Dans notre étude, la réduction progressive de la loss et l'augmentation de l'accuracy au fil

des EPOCH montrent que le modèle apprend efficacement et devient de plus en plus performant.

L'analyse des spectrogrammes a révélé que certains sons, tels que les coups de feu (gun_shot) et les sirènes (siren), sont facilement distinguables grâce à des caractéristiques spectrales distinctes. Cependant, d'autres sons, comme ceux des enfants jouant (children_playing) et les sirènes, peuvent parfois être confondus en raison de similitudes dans leurs signatures spectrales. Ces observations soulignent l'importance d'une extraction précise des caractéristiques et d'un prétraitement rigoureux pour améliorer la différenciation des sons similaires.

Voici quelques exemples des spectrogrammes obtenus :

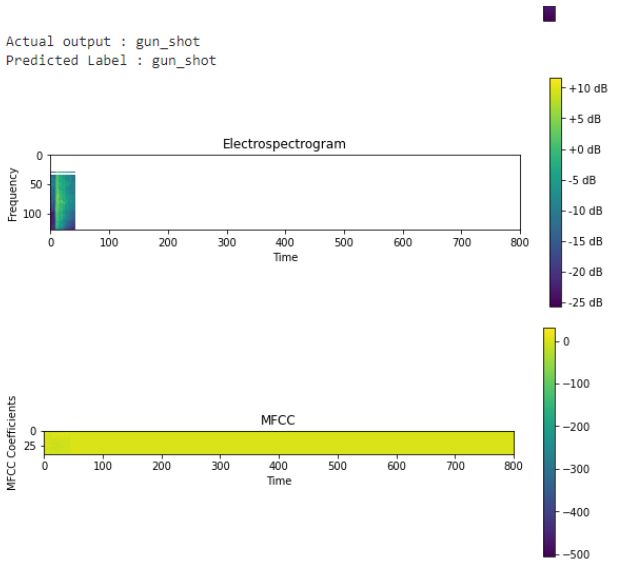

Figure 2: Gun shot

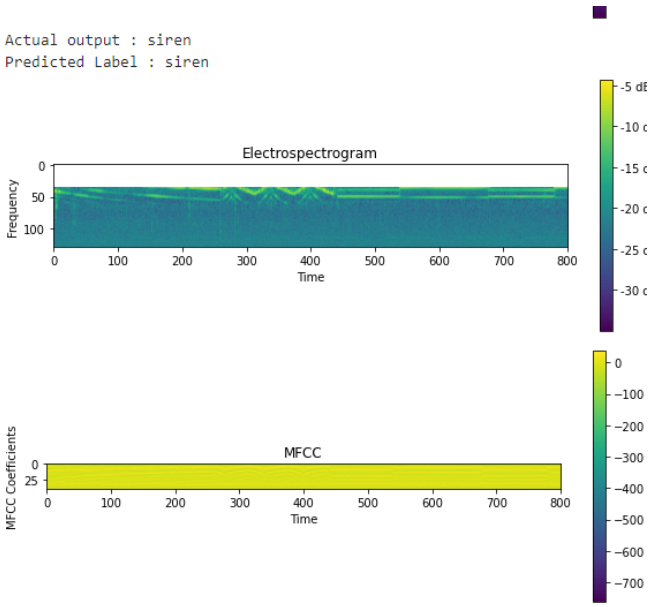

Figure 3: Siren

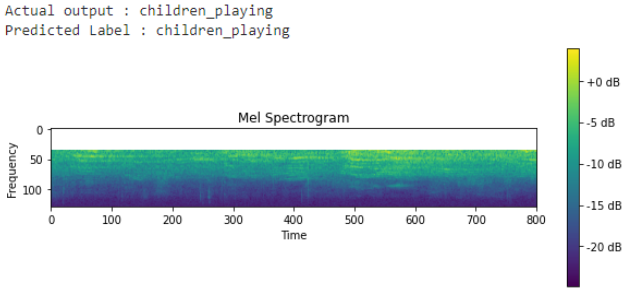

Figure 4: Children playing

Ces spectrogrammes montrent que les sons comme les coups de feu et les sirènes peuvent être clairement identifiés grâce à des caractéristiques spectrales distinctes, tandis que d'autres sons peuvent présenter des défis en raison de leur similitude.

Les hypothèses formulées dans cette étude mettent en évidence plusieurs aspects cruciaux de l'utilisation des réseaux de neurones convolutifs (CNN) pour la classification des sons d'oiseaux rares et nocturnes. Tout d'abord, il est postulé que les CNN sont particulièrement efficaces pour cette tâche, même avec des jeux de données limités, en montrant une amélioration continue des performances (précision, rappel, F1-score) dès les premières EPOCH. Ensuite, le prétraitement des données, incluant la réduction du bruit et la normalisation, est jugé essentiel pour améliorer significativement les performances du modèle. Ces étapes augmentent la clarté et la qualité des enregistrements audio, conduisant à une diminution de la perte (loss) et à une augmentation de la précision (accuracy). De plus, les techniques de régularisation, telles que le dropout et la normalisation de lot, sont considérées cruciales pour stabiliser les performances du modèle et éviter le surapprentissage, permettant ainsi au modèle de généraliser efficacement sur des données non vues.

Enfin, l'utilisation de techniques de data augmentation est hypothétisée pour enrichir le jeu de données et améliorer la capacité de généralisation du modèle, se traduisant par des performances stables et élevées en termes de précision et de rappel.

L'analyse des spectrogrammes des sons d'oiseaux a révélé des motifs distincts correspondant aux caractéristiques uniques des chants d'oiseaux. Par exemple, les spectrogrammes des enregistrements de l'Asian Brown Flycatcher et de l'Ashy Drongo montrent des bandes de fréquence spécifiques et des motifs acoustiques distincts.

En revanche, les spectrogrammes des sons urbains ont montré une plus grande variabilité et complexité due aux multiples sources de bruit présentes dans ces environnements, tels que les sirènes, les klaxons et les moteurs. (voir graphiques ci-dessus)

L'analyse comparée des spectrogrammes des deux espèces d'oiseaux montre des différences significatives dans leurs chants, tant en termes de fréquences dominantes que de motifs acoustiques et de distribution temporelle. Le spectrogramme de l'Asian Brown Flycatcher révèle des bandes de

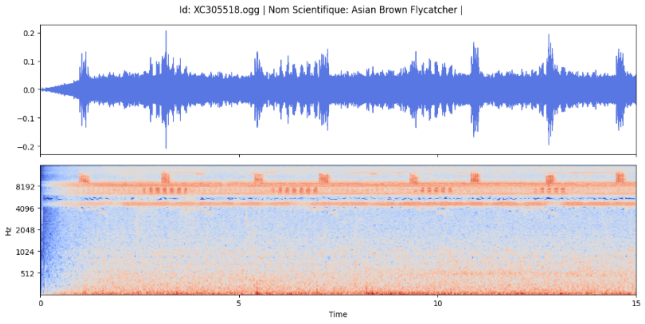

Figure 5: Asian Brown

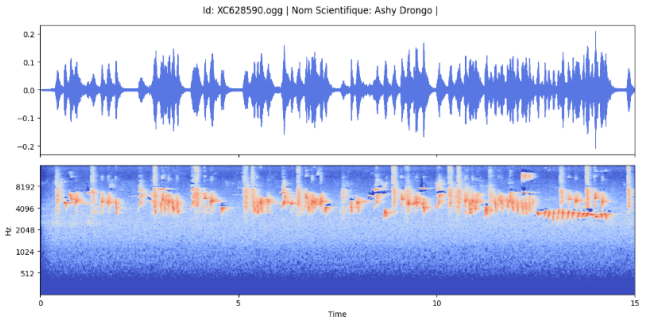

Figure 6: Ashy Drongo

fréquence distinctes et répétitives autour de 1000 à 4000 Hz, avec des pauses plus régulières et des motifs plus espacés, indiquant une communication structurée et intermittente. En revanche, le spectrogramme de l'Ashy Drongo montre une activité fréquente dans les bandes de fréquence autour de 2000 à 6000 Hz, avec des motifs plus complexes et une densité de points élevés, suggérant un chant plus varié et dynamique.

2.2 Ouverture

Nos résultats montrent que, bien que les sons urbains puissent parfois masquer les chants d'oiseaux, les techniques de prétraitement des données et d'extraction des caractéristiques permettent aux CNN de distinguer efficacement ces deux types de sons. Cette recherche valide l'utilisation des CNN pour la classification des sons d'oiseaux rares et nocturnes et souligne l'importance du prétraitement des données pour améliorer la précision et la robustesse des modèles de machine learning.

L'étude ouvre plusieurs perspectives intéressantes pour les recherches futures et les applications pratiques. Bien que les CNN aient démontré une grande efficacité pour la classification des sons d'oiseaux rares et nocturnes, d'autres techniques de machine learning telles que les forêts aléatoires, les machines à vecteurs de support (SVM) et les modèles basés sur les transformations spectrales pourraient également être explorées pour améliorer les performances de classification. Des approches hybrides combinant des CNN avec d'autres modèles de machine learning pourraient offrir de meilleures performances.

De plus, l'augmentation des données d'entraînement par des techniques de data augmentation et la synthèse de données réalistes pourraient enrichir le jeu de données et améliorer la robustesse du modèle. L'intégration de données provenant de multiples sources, y compris des enregistrements de terrain et des bases de données publiques, pourrait capturer une plus grande diversité de sons et de conditions environnementales, augmentant ainsi la généralisation du modèle.

Les résultats de cette étude peuvent être appliqués dans des domaines tels que la surveillance de la biodiversité, la conservation des espèces rares et nocturnes, et la gestion des écosystèmes, en facilitant la détection et la classification automatiques des sons d'oiseaux dans des environnements naturels.

2.3 Références

"Large-Scale Bird Sound Classification using Convolutional Neural Networks"

- **Auteurs :** Stefan Kahl, Thomas Wilhelm-Stein, Hussein Hussein, Holger Klinck, Danny Kowerko, Marc Ritter, Maximilian Eibl
- **Journal :** CEUR Workshop Proceedings, 2017

"Investigation of Different CNN-Based Models for Improved Bird Sound Classification"

- **Auteurs :** J. Xie, S. Koch, T. Kullback, H. Klinck
- **Journal :** IEEE Access, 2021

"A CNN Sound Classification Mechanism Using Data Augmentation"

- **Auteurs :** Lin Wang, Xiaofeng Zhou, Huihui Wang
- **Journal :** MDPI Sensors, 2020

"Handcrafted features and late fusion with deep learning for bird sound classification"

- **Auteurs :** B. LeBien, S. Williams, C. J. Trainor
- **Journal :** Ecological Informatics, 2020

"Bird Sound Recognition Using a Convolutional Neural Network"

- **Auteurs :** M. Gerhard, T. Schindler, K. Römer
- **Journal :** IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2020