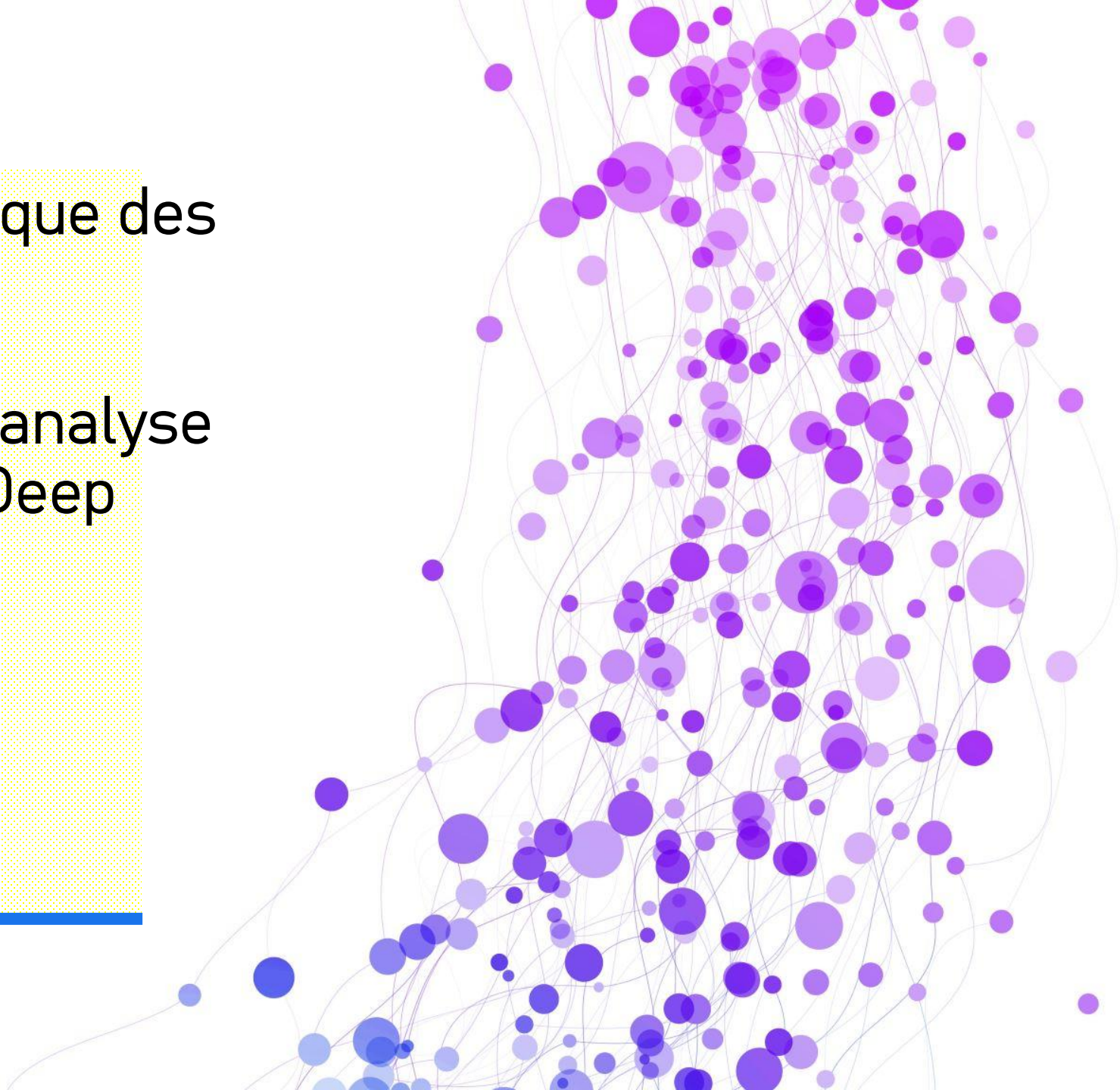


Classification acoustique des
drones :

Approche combinant analyse
temps-fréquence et Deep
Learning

PROJET SYSTÈME - SOIA 2025

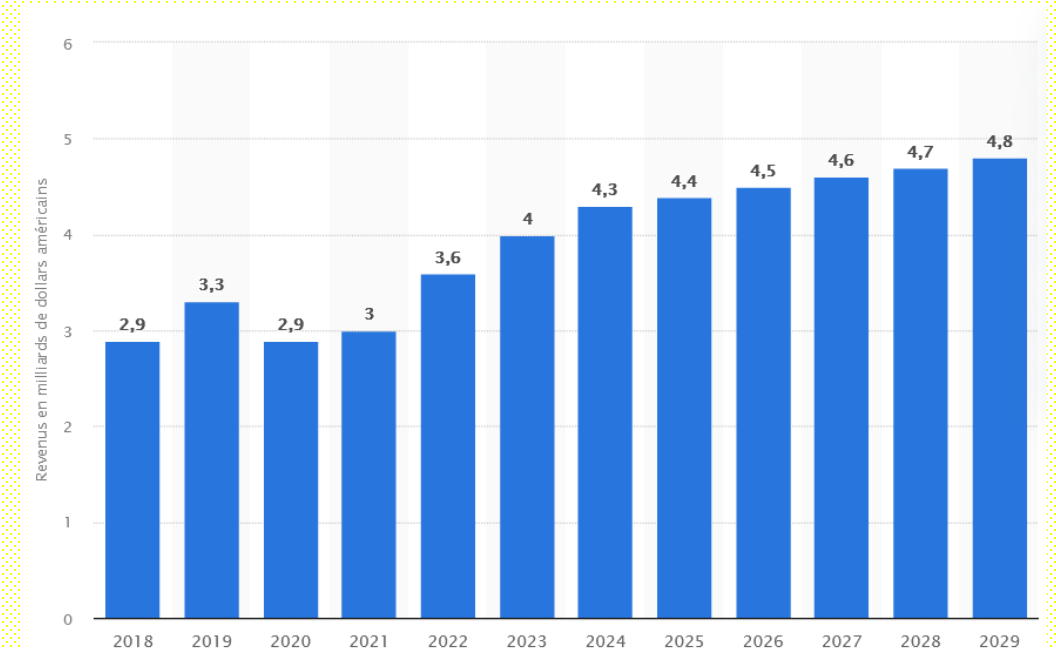
RAPHAËL GRISEL



Contexte

En France :

- 103 000 drones vendus en 2020
- Parc de drones de loisirs en France estimé à 2,5 millions d'appareils



Marché mondial du drone grand public [1]

[1] Maxime Gautier, 14 janv. 2025 , <https://fr.statista.com/statistiques/607544/croissance-prevue-marche-drones-professionnels-monde/>

Contexte



Attaque sur le président Venezuelien Maduro en 2018 [2]



Crash de drone imprévisible

[2] T. Guardian, "P. daniels, venezuela's nicolas maduro survives apparent assassination attempt," <https://www.theguardian.com/world/2018/aug/04/nicolas-maduros-speech-cut-short-while-soldiers-scatter>., 2018.

Sommaire

- Etat de l'art:
 - Méthodes de classification
 - Méthodes TFT et Deep Learning
- Projet
 - Le Dataset
 - Le Modèle
 - Détection de drones
 - Classification
 - Représentation en spectrogramme de Mel
- Retour sur le projet

Etat de l'art : méthodes de machine learning

Détection et classification de drones

Système d'observation :

- Acoustique
- Radar
- Vision
- Radiofréquence

Modèle de machine Learning :

- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)
- Recurent Convolutional Neural Network (RCNN)
- Support Vector Machine (SVM)

Etat de l'art : méthodes de machine learning

Méthode	Avantages	Inconvénients
Acoustique	Etude de la signature sonore Méthode peu intrusive	Portée de détection limitée Sensible au bruit ambiant
Radar	Longue portée Peu sensible aux conditions météorologiques	Pas adapté aux drones de faible SER
Vision	Caméras peu coûteuses Fournit des informations visuelles	Dépend de la ligne de mire Sensible aux conditions météorologiques
Radiofréquence	Localiser les contrôleurs du drone. Longue portée	Inutile si le drone est en mode autonome Sensible au brouillage et aux interférences

Acquisition de données pour la détection et la classification de drones [3]

Etat de l'art:

Méthode acoustique :

- Peu coûteux
- Peu technique

MAIS

- Manque de données d'entraînement
- Prise en compte du bruit ambiant

Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks

Jun Lee, Woong-Hee Kim, YoungHyouon Kwon, and Hae-Yong Yang
The Affiliated Institute of ETRI
Daejeon, South Korea
{junlee, whkim, wshwill, formant}@etri.re.kr

Audio Based Drone Detection and Identification using Deep Learning

Sara Al Emadi, Abdulla Al Ali, Amr Mohammad, Abdulaziz Al Ali
Department of Computer Science and Engineering
Qatar University
Doha, Qatar
saraalemadi@ieee.org, {abdulla.alali, amr, a.ali}@qu.edu.qa

Abstract—In recent years, unmanned aerial vehicles (UAVs) have become increasingly accessible to the public due to their high availability with affordable prices while being equipped with better technology. However, this raises a great concern from both the cyber and physical security perspectives since UAVs can be utilized for malicious activities in order to exploit vulnerabilities by spying on private properties, critical areas or to carry dangerous objects such as explosives which makes them a great threat to the society. Drone identification is considered the first step in a multi-procedural process in securing physical infrastructure against this threat. In this paper, we present drone detection and identification methods using deep learning techniques such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Convolutional Recurrent Neural Network (CRNN). These algorithms will be utilized to exploit the unique acoustic fingerprints of the flying drones in order to detect and identify them. We propose a comparison between the performance of different neural networks based on our dataset which features audio recorded samples of drone activities. The major contribution of our work is to validate the usage of these methodologies of drone detection and identification in real life scenarios and to provide a robust comparison of the performance between different deep neural network algorithms for this application. In addition, we are releasing the dataset of drone audio clips for the research community for further analysis.

Index Terms—Drone, UAVs, Acoustic fingerprinting, Drone Audio Dataset, Artificial Intelligence, Machine Learning

I. INTRODUCTION AND RELATED WORK

Due to the development of drone technology, the popularity of drones is rapidly increasing as they are becoming more compact in size, easier to operate and widely available for anyone with a desire to use them. Many utilize these drones for recreational purposes such as photography and cinematography, others are studying their behavior and improving on them in order to integrate them in our day to day applications. On the other hand, it has been observed that drones can be utilized for malicious activities to harm targeted individuals and the public. Recently, an incident was reported to authorities in which an explosive equipped drone was hovering over a great crowd in a formal occasion in Venezuela targeting a high profile personnel and the general public. In this incident the drone dropped a number of attached explosives randomly which consequently injured civilians on scene [1]. In addition to the safety issues associated with drones, they can be used to violate security terms and conditions as it has been witnessed in an incident in which smugglers flew drones with illegal

drugs and cell phones over prison facilities [2]. Similarly, privacy concerns arose with the introduction of drones where at multiple incidents they were used to spy and record clips of people in their private properties [3][4][5].

As with any device, unintentional accidents beyond the drone operator's control can occur. For example, the connection between the controller and the drone can be lost causing the drone to drift away from the predefined path and crash into its surrounding causing serious damages. This scenario occurred recently in an incident where a drone have severely injured a toddler after crashing into a children's playground [6].

A key objective of this paper is to introduce an autonomous system that in addition to *detection*, is able to *identify* drones based on their acoustic signatures. The identification of the drones is necessary in order to determine the type of the drone which will aid in the process of distinguishing whether the drone flying within a restricted area is an authorized or an unauthorized drone.

There is now a substantial body of research on the application of drone detection using different technologies such as the RF signals [7] [8], a GSM passive coherent location system [9] and a digital TV based bi-static radar [10]. Furthermore, few researchers focused their studies on drone detection using audio characteristics such as a research carried out by the authors in [11], where they have proposed a methodology using digital signal processing (DSP) to detect the presence of drones in an area. Another research serving the same purpose was conducted in [12], in which authors opted for combining DSP with Machine Learning algorithms such as the Support Vector Machine (SVM) algorithm. The researchers have reported the effectiveness of using SVM in drone detection which have yielded high accuracy, yet, the research was limited to a specific background noises. Furthermore, SVM requires manual extraction and optimization of hand-crafted features to fine tune the algorithm, this is an additional step to the actual classification problem. However, using deep learning models will eliminate this issue by ensuring an end to end training of the model autonomously [13]. Similar approach was put forward by the authors in [14] to target drone detection using DSP along with two Machine Learning algorithms, the Plot Image Learning (PIL) and the K-Nearest Neighbor (KNN). Although the detection ability proved its effectiveness, the

use of deep learning models in real-life environments for malicious activities by drones as a detection method is still a challenge. Our empirical study collected on these models, our RNN is an F-Score of 0.95, and our CNN is 0.92.

hobby drones which we live, a common hobby and for singly makes (acy systems, is can easily ch as landing ie rooftop of an, and at the lity to detect (best priority

ve been con- ion, previous t rather than : sound envi- rooftop of a [3], [2], [3], t is to detect t this requires rk differs by formation or r a combined

detection system with a multiple approach to complement the drawback of each method.

Event Sound Classification (ESC) in a real environment has been highlighted for diverse purposes. Many researchers have focused on finding useful features and classifiers based on the machine-learning approach. The most popular combination of feature and classification is Mel-frequency Cepstrum Coefficients (MFCC) [15] with the Gaussian Mixture Model (GMM) [16], [17]. More recently, the impressive success achieved with Deep Neural Networks (DNNs) has motivated researchers to introduce these networks to environmental sound recognition. Two popular DNN models, the Convolutional Neural Network (CNN) [18], [19] and Recurrent Neural Network (RNN) [20], [21], have also been highlighted for audio-related tasks. Even though these previous studies cover the ESC problem, considering the importance and urgency of our problem in terms of terrorism, it is worth exploring how ESC work can be applied and to assess its effectiveness for drone sound detection. Here it should be noted that rather than intended to propose novel features or models for drone sound detection, our work aims to investigate the practical effectiveness of popular classification models for our problem in real environments used in previous ESC studies.

Contribution. Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to investigate drone sound detection in highly noisy real environments with the aim of constructing a detection method for practical usage with real-time systems based on three popular ESC models: GMM, CNN, and RNN.
- We show that the shortage of training data for a drone sound classification model can be remedied with our audio augmentation that synthesizes raw drone sound with diverse background sounds.
- We investigate the effectiveness of these models for a testing dataset collected from real life environments in terms of the F-Score and by taking consideration of the processing time for application to real-time systems.

II. METHOD

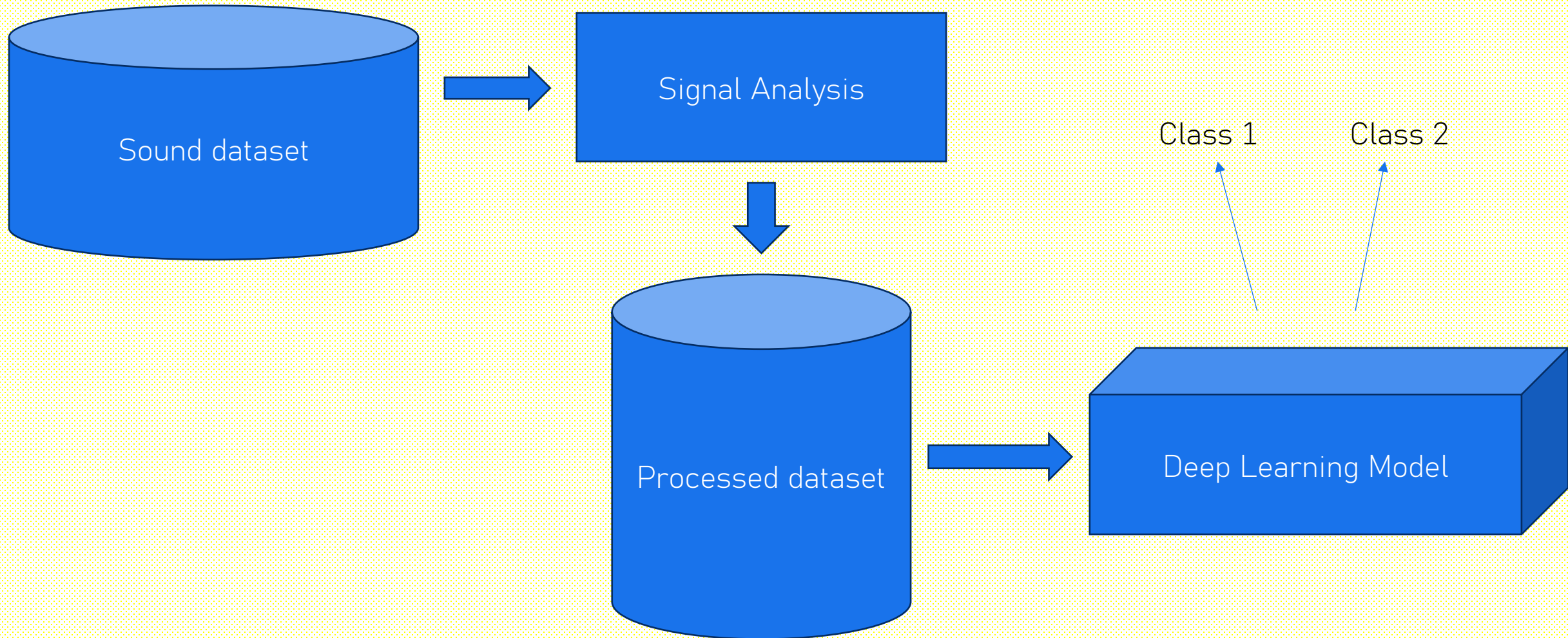
A. Data Augmentation

Especially in real environments, unseen event sound has a detrimental effect in terms of deterioration of the detection rate.

[4] A. M. A. A.-A. Sara Al-Emadi, Abdulla Al-Ali, "Audio based drone detection and identification using deep learning," 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), 2019

[5] Sungho Jeon¹, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyouon Kwon, and Hae-Yong Yang "Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks", 2017arXiv

Etat de l'art : chaîne de traitement



Etat de l'art : analyse temps fréquence

Représentation conjointe du signal dans les domaines temporel et fréquentiel

- Analyse des signaux non stationnaires
- Extraction des caractéristiques pertinentes
- Classification basée sur l'imagerie

Etat de l'art :

Mis à disposition d'un dataset d'enregistrement de drones

Détection et classification sur des spectrogrammes d'enregistrement sonores de drones

Audio Based Drone Detection and Identification using Deep Learning

Sara Al Emadi, Abdulla Al Ali, Amr Mohammad, Abdulaziz Al Ali
Department of Computer Science and Engineering
Qatar University
Doha, Qatar
saraalemadi@ieee.org, [abdulla.alali, amrm, a.alali]@qu.edu.qa

Accuracy des modèles pour les tâches de détection et classification [4]

	Détection	Classification
CNN	96,38 %	92,94%
RNN	75,00%	57,16%
CRNN	94,72%	92,22%

CNN est le modèle le plus performant pour la détection et la classification

4] A. M. A. A.-A. Sara Al-Emadi, Abdulla Al-Ali, "Audio based drone detection and identification using deep learning," 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC),2019

Objectifs du projet

- Implémenter un algorithme de détection et de classification acoustiques sur le dataset proposé par [4]
- Etudier l'influence de différentes Transformations Temps-Fréquence sur les performances du modèle de Deep Learning

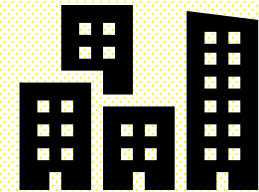
Le Dataset



Bebop [6]



Mambo [6]

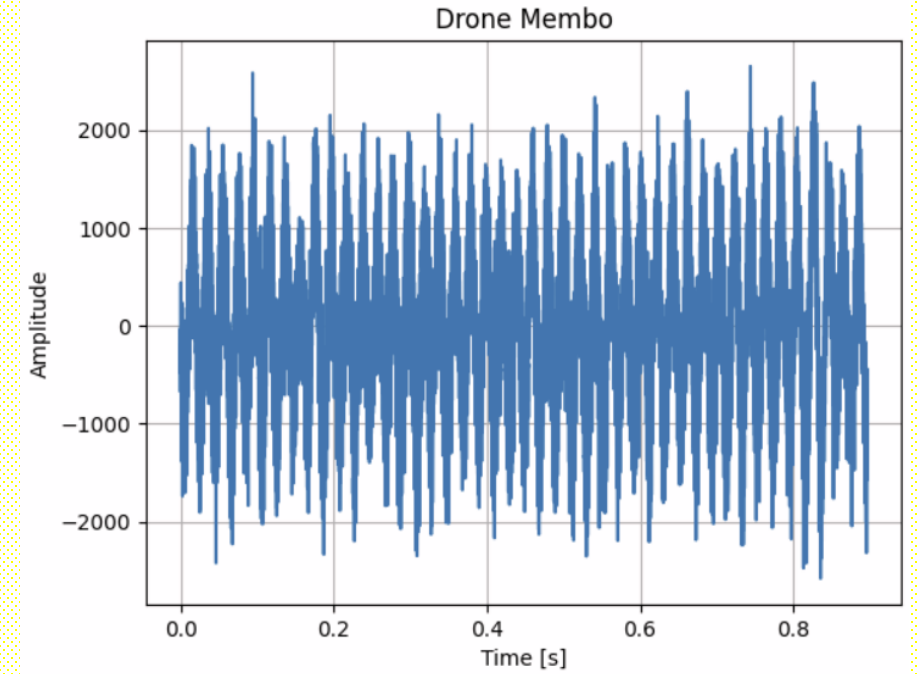
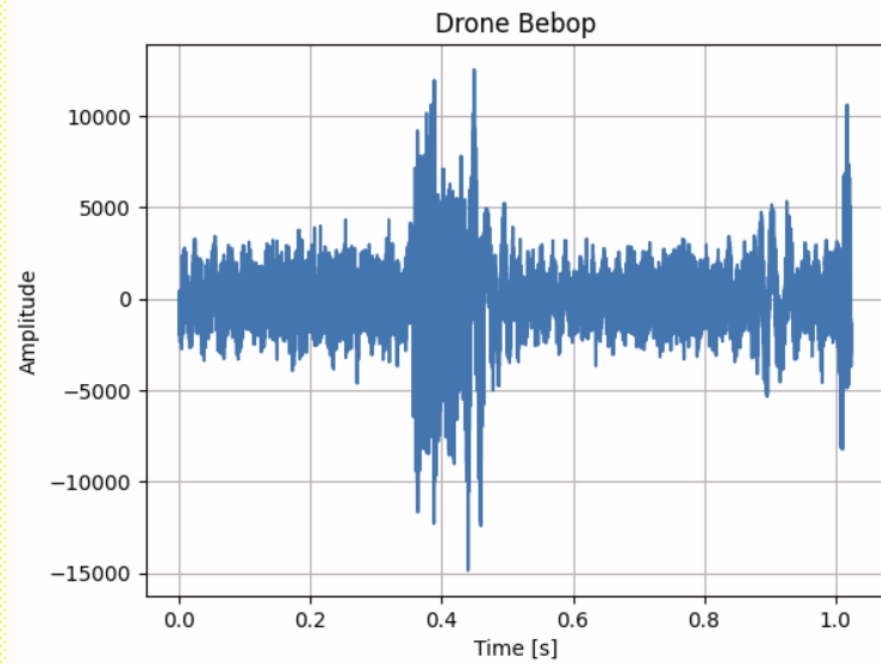


Environnement
sonore

Le Dataset

- ❖ Acquisition :
 - ❖ microphone d'un smartphone
 - ❖ environnement intérieur calme
 - ❖ 11 minutes et 6 seconds par drone
- ❖ Augmentation :
 - ❖ Simuler enregistrements réalistes
 - ❖ Sons drones + sons environnements
- ❖ Mise en forme :
 - ❖ Etiquetage
 - ❖ Découpage en segment de 1 seconde
 - ❖ Dataset Détection : 2 classes
 - ❖ Dataset Classification : 3 classes

Analyse des enregistrements

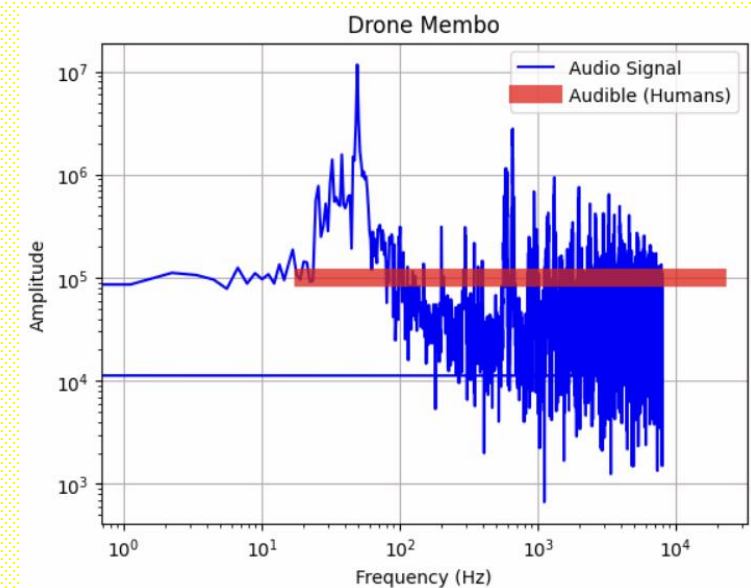
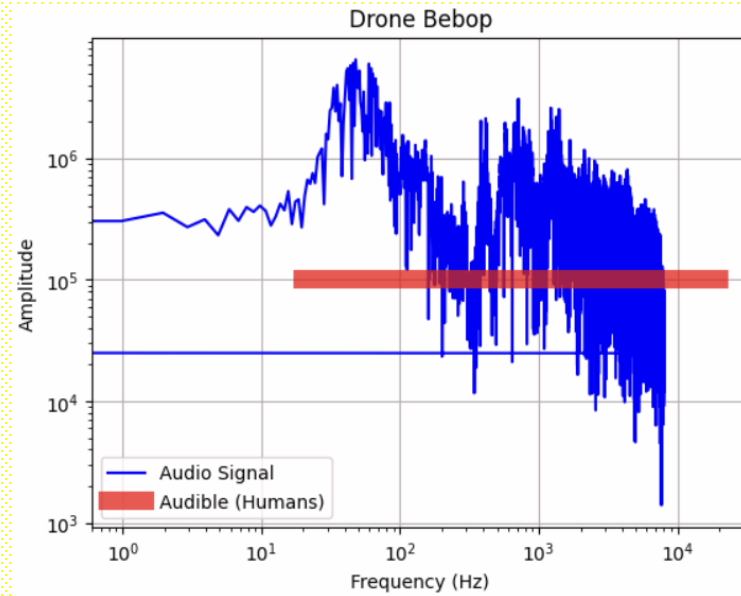


Visualisations temporelles des enregistrements sonores

Analyse des enregistrements

Fast Fourier Transform :

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi kn}{N}}, \quad k = 0, 1, 2, \dots, N-1$$



Analyse des enregistrements

❖ Spectrogrammes :

❖ Transformée de Fourier à court terme:

$$X(m, k) = \sum_{n=0}^{N-1} x[n]w[n - mR]e^{-j\frac{2\pi kn}{N}}$$

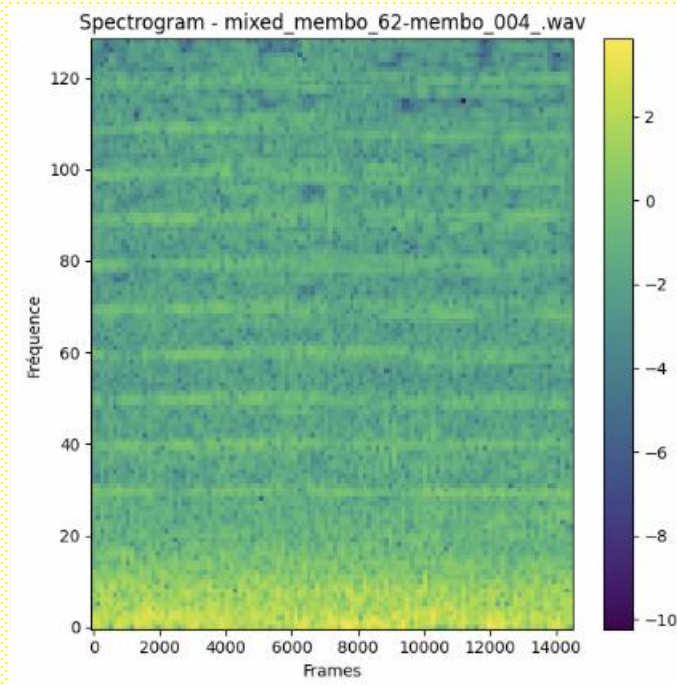
❖ Spectre en amplitude :

$$S(m, k) = |X(m, k)|^2$$

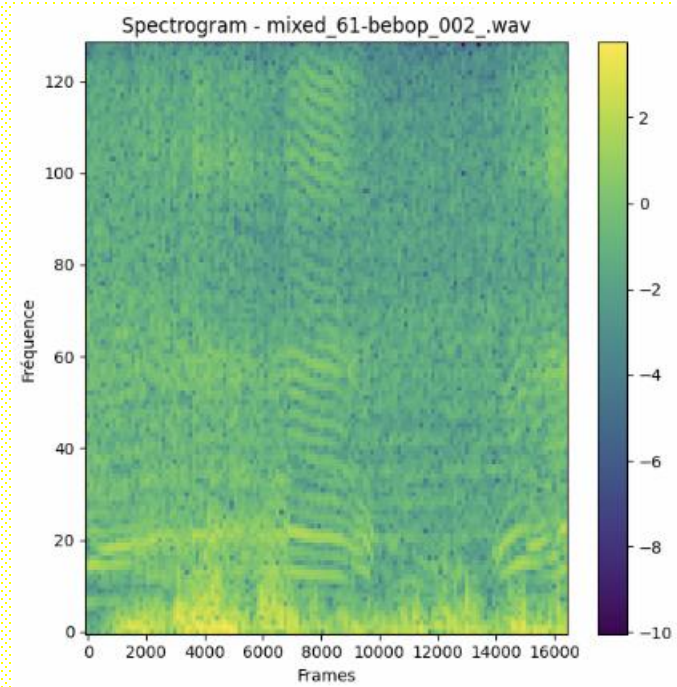
❖ Conversion en dB :

$$S_{dB}(m, k) = 10 \log_{10} S(m, k)$$

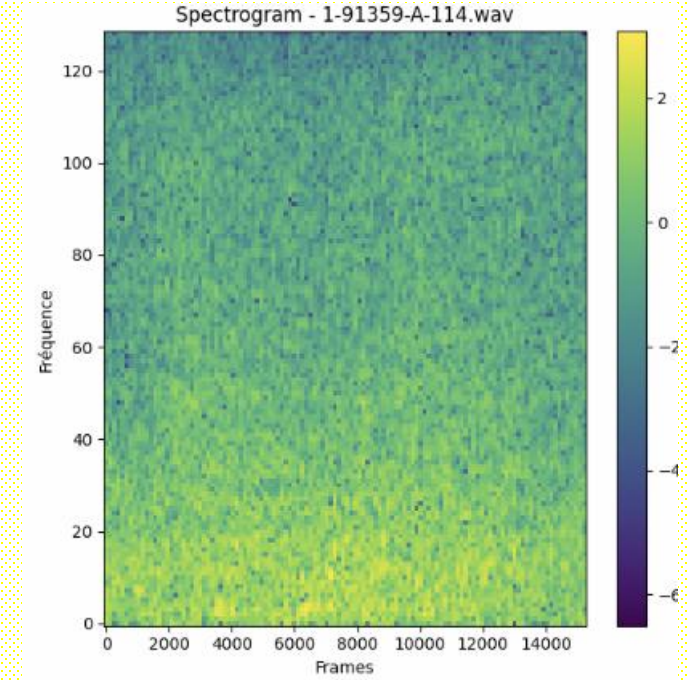
Analyse des enregistrements



Mambo

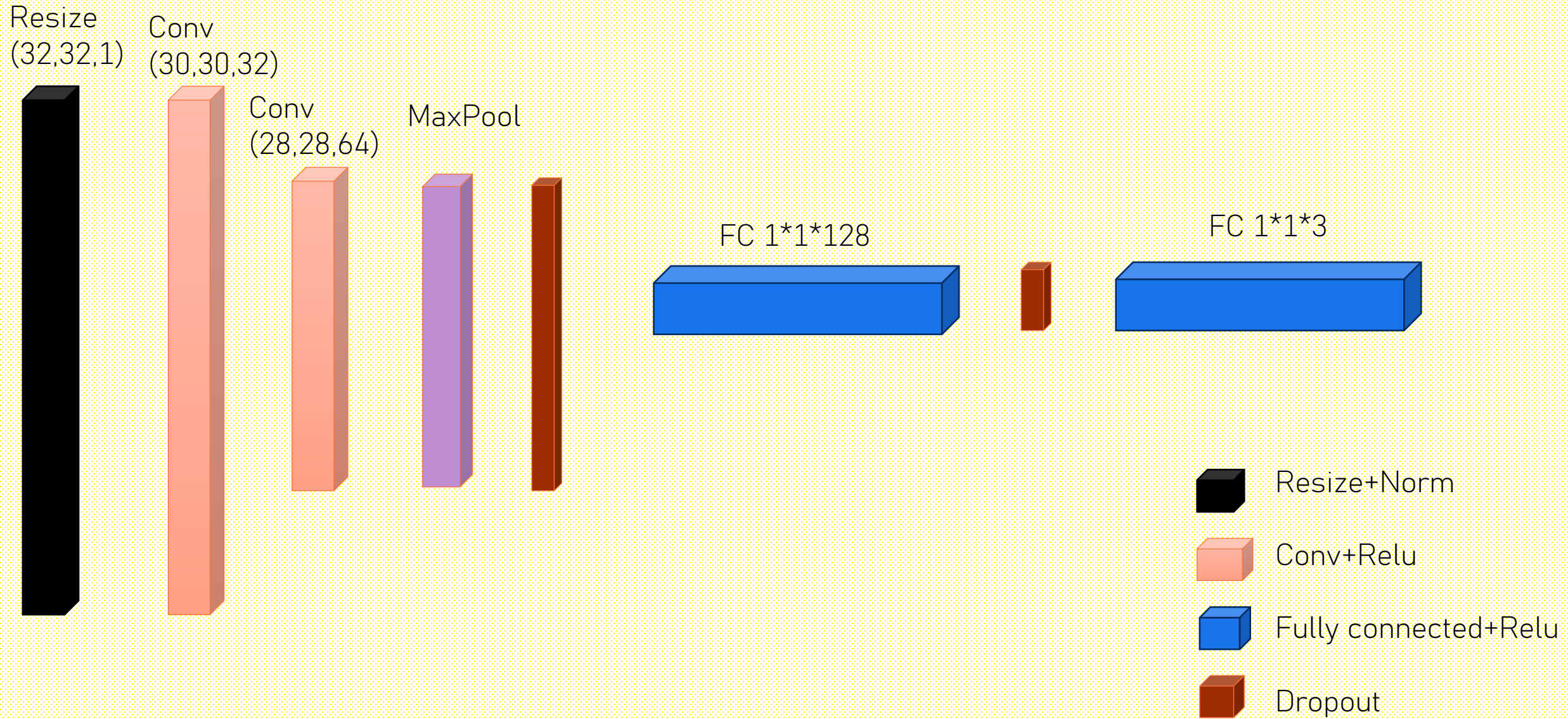


Bebop



Environnement

Modèle de Deep Learning : CNN



Modèle de Deep Learning : CNN

Processus d'entraînement :

- ❖ Optimiseur : Adam
- ❖ Learning rate : Exponential Decay Scheduler
- ❖ Loss : Sparse Categorical Crossentropy :

$$L = -\frac{1}{N} \sum_{i=1}^N \log P(y_i)$$

- ❖ Early stopping : après 5 epochs consecutive sans amélioration de la loss
- ❖ Entraînement sur 50 epochs

Modèle de Deep Learning : CNN

Processus d'évaluation :

❖ Accuracy :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

❖ F1-score:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

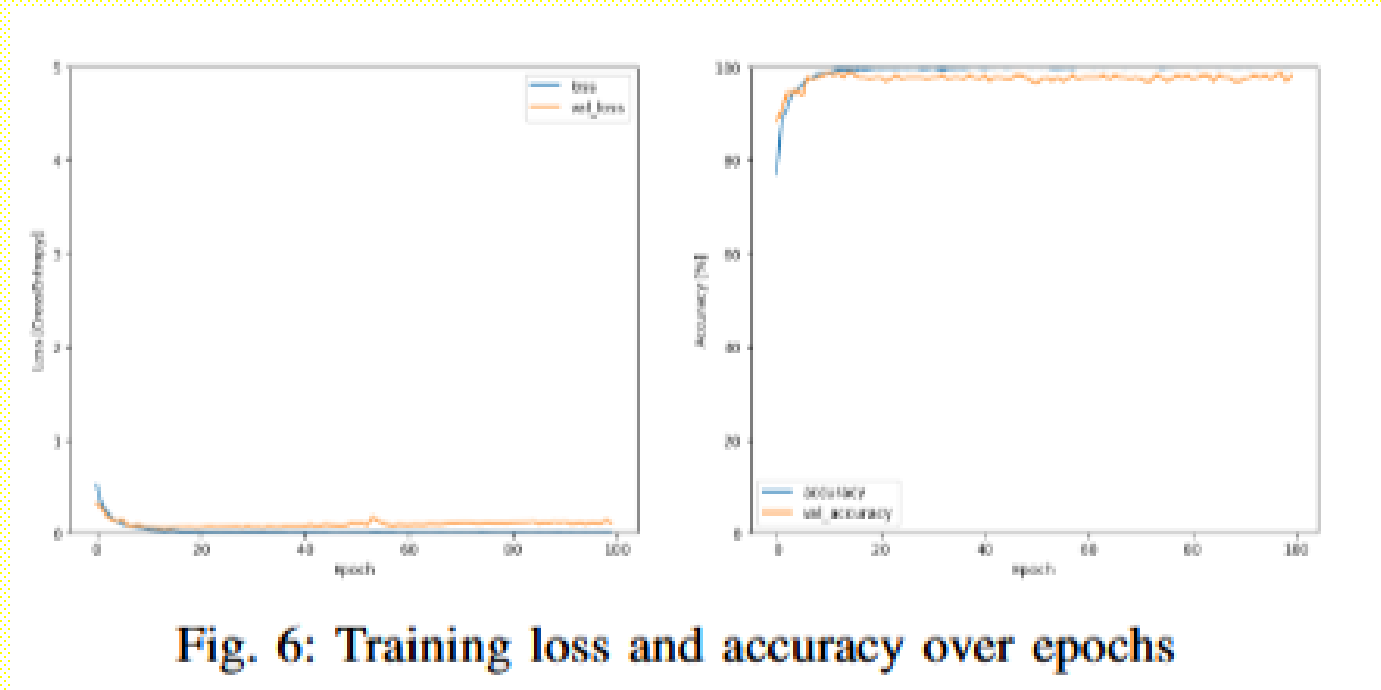
❖ Matrice de confusion:

Performances évaluées sur 10
entraînements de 50 epochs

Détection

Entraînement :

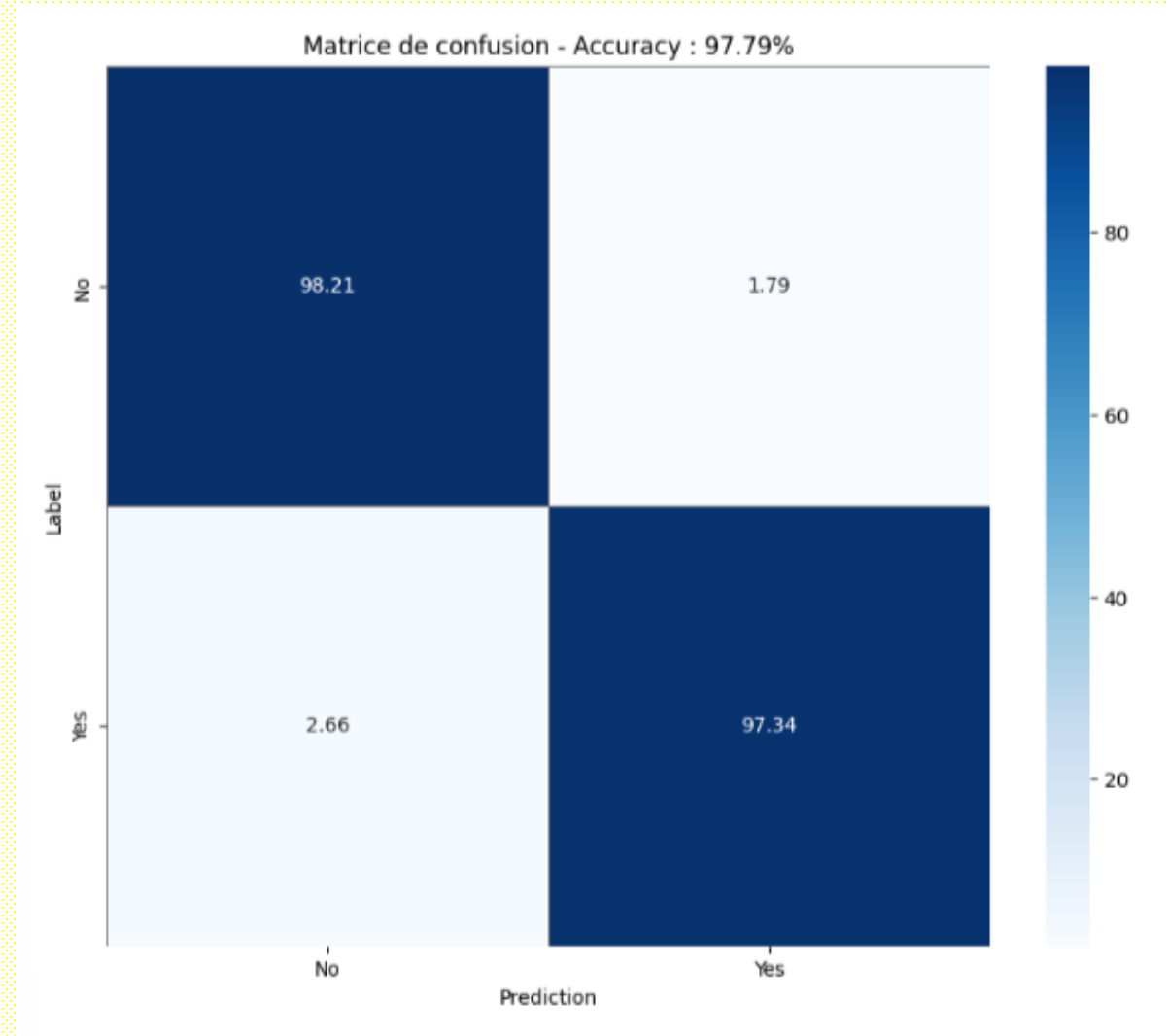
- Avec scheduler et early stopping : convergence du modèle



Détection

Evaluation du modèle:

- Accuracy : 97,62%
- F1-score : 95,90%

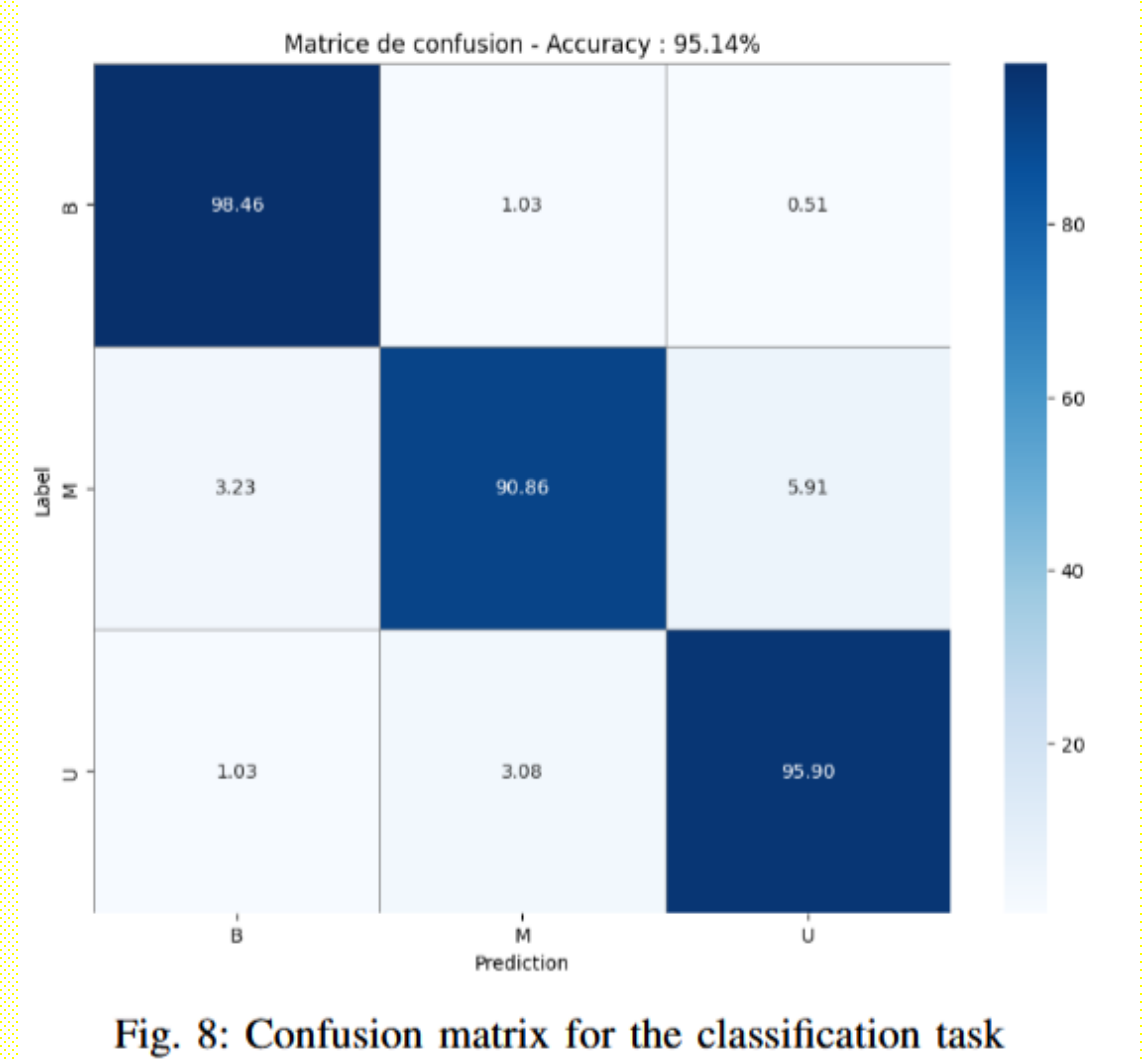


Meilleures performances que le modèle de détection du papier [4]

Classification multiclass

Evaluation du modèle:

- Accuracy : 98,46%
- F1-score : 95,90%



Meilleures performances que le modèle de classification du papier [4]

Interprétation des résultats

	Detection	Classification
Our model	97,62%	98,46%
Paper's model [4]	96,38%	92,94%

Amélioration significative de l'accuracy pour la classification multiclass

- ❖ Equilibrage des classes
- ❖ Normalisation des spectrogrammes
- ❖ Choix de l'architecture du model (Dropout, Conv, Dense)
- ❖ Choix des hyperparamètres (LR, Loss, EearlyStopping)

Spectrogrammes de Mel

Représentation proche des perceptions humaines , diminue les détails des hautes fréquences

Spectrogrammes Mel :

❖ Transformée de Fourier à court terme:

$$X(m, k) = \sum_{n=0}^{N-1} x[n]w[n - mR]e^{-j\frac{2\pi kn}{N}}$$

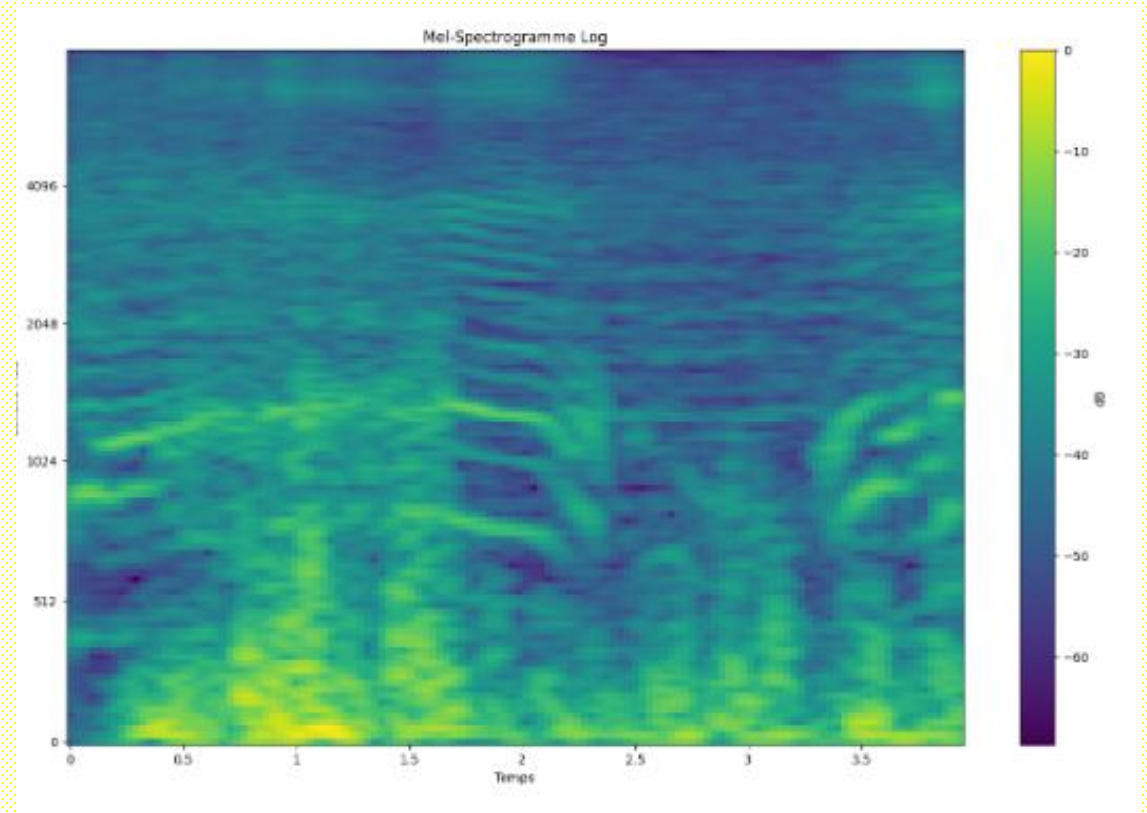
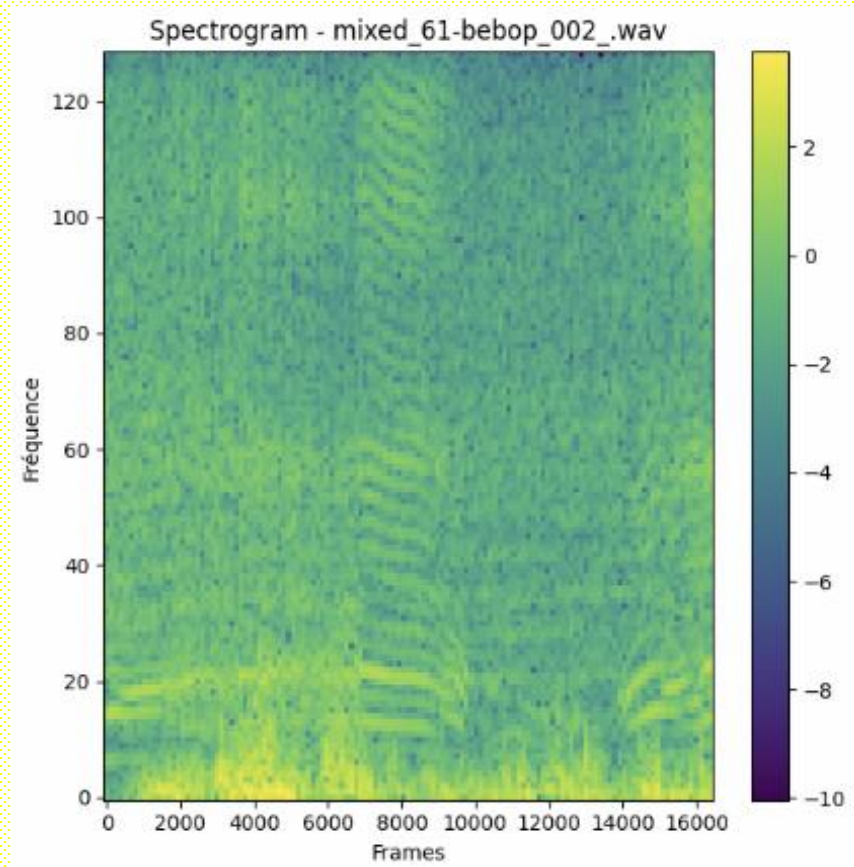
❖ Spectre en amplitude :

$$S(m, k) = |X(m, k)|^2$$

❖ Conversion des fréquences linéaire en échelle de Mel:

$$f_{\text{mel}} = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

Spectrogrammes de Mel



Spectrogramme et Spectrogramme de Mel de Bebop

Détection et Classification sur spectrogrammes Mel

	Détection	Classification
Spec + CNN	97,62%	98,46%
Mel-Spec + CNN	97,14%	87,4%

Chute significative de performances pour la classification avec les spectrogrammes de Mel

Interprétation des résultats

- ❖ Contenu des plus hautes fréquences différencie les drones entre eux
- ❖ Contenu des plus basses fréquences discrimine les drones et l'environnement
- ❖ Filtre de Mel adoucie la représentation fréquence : perte de détails

Retour sur le projet

- ❖ Implémentation d'un algorithme adaptable de détection et de classification acoustique de drones
- ❖ Modèle plus performant que le papier de référence [4]
- ❖ Etudier les spécificités des représentation et Spectrogrammes et spectrogrammes de Mel et leur impacte sur les performances du modèle

Conclusion

- Méthode combinant représentation Temps-fréquence et Deep Learning montre des bons résultats
- Le choix de la TFT impacte les performances du modèle
- Faible coût , peu intrusive, modèle performant : classification acoustique combinant TFT et deep learning est une solution efficace aux problématiques soulevées par l'essor des drones de loisir

Références Bibliographiques

- [1] Maxime Gautier, 14 janv. 2025 , <https://fr.statista.com/statistiques/607544/croissance-prevue-marche-drones-professionnels-monde/>
- [2] T. Guardian, "P. daniels, venezuela's nicolas maduro survives apparent assassination attempt," <https://www.theguardian.com/world/2018/aug/04/nicolas-maduros-speech-cut-short-while-soldiers-scatter.>, 2018.
- [3] B. TAHA and A. SHOUFAN, "Machine learning-based drone detection and classification: State-of-the-art in research," IEE, 2019.
- [4] A. M. A. A.-A. Sara Al-Emadi, Abdulla Al-Ali, "Audio based drone detection and identification using deep learning," 2019 15th International Wireless Communications Mobile Computing Conference (IWCMC),2019
- [5] Sungho Jeon¹, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyouun Kwon, and Hae-Yong Yang "Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks", 2017arXiv

Références Bibliographiques

- [6] parrot.com
- [7] V. C. Georgopoulos, "Advanced time-frequency analysis and machine learning for pathological voice detection," 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), 2020.
- [8] R. B. Constantin CONSTANTINESCU, "An overview on sound features in time and frequency domain," Sciendo, 2023.

Annexe

