

Compte Rendu du Projet : Classification et Détection Audio

Loïs Goncet

I. Détection des Notes et Traitement du Signal

A. Détection de la note

Pour déterminer la note jouée à un instant donné, une approche par analyse fréquentielle a été employée

- **Méthode** : Le signal audio est analysé par segments de **0,25 seconde**. Une **Transformée de Fourier Rapide (FFT)** est appliquée à chaque segment pour obtenir son spectre. La fréquence dominante est ensuite identifiée et convertie en note de musique (ex: Sol4, Do5).
- **Implémentation** : Cette logique a été intégrée pour retourner la séquence des notes jouées pour un instrument unique, ce qui répond à l'objectif initial.

B. Construction des Spectrogrammes (Pré-traitement)

Le pré-traitement des données est l'étape la plus critique. Pour traiter l'audio comme une image, nous utilisons des spectrogrammes.

- **Spectrogrammes Multi-Fenêtres** : Nous avons adopté une approche avancée en calculant **trois spectrogrammes différents** pour chaque fichier audio, en variant les paramètres de la fenêtre d'analyse (Hamming).
 - Chaque spectrogramme (Fenêtre étroite, moyenne, large) met en évidence différents aspects du son (transitoires, fréquences moyennes, basses fréquences).
 - Ces trois vues sont empilées pour former une image **3D à 3 canaux** (similaire au format RGB), offrant au modèle une information riche et complète.
- **Redimensionnement** : Les images ont été redimensionnées à **(128, 128, 3)** pour optimiser la vitesse d'entraînement sur le GPU.
- **Format de Stockage** : Les spectrogrammes ont été sauvegardés en format **.npy** (NumPy array) pour préserver la précision des nombres flottants.

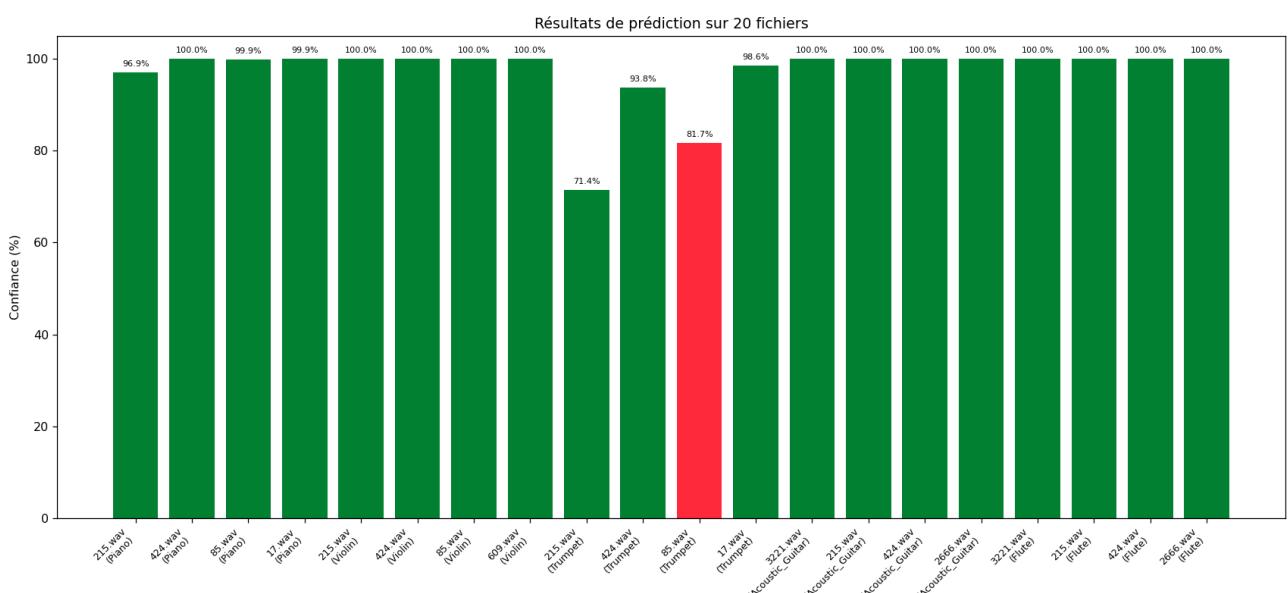
II. Détection d'un Unique Instrument (Classification Simple)

A. Entraînement et Modèle

- **Architecture** : Le modèle choisi est le **ResNet-34** (Réseau de Neurones Résiduel à 34 couches), implémenté et entraîné **depuis zéro** (from scratch) pour classer les 26 instruments disponibles.
- **Stratégie Anti-Overfitting** : L'apprentissage initial était instable. Pour assurer la robustesse du modèle (généralisation au lieu de mémorisation), deux techniques ont été mises en place :
 1. **Augmentation des Données** (zoom et retournement aléatoires) pour diversifier les images d'entraînement.
 2. **Dropout** (désactivation aléatoire de neurones) ajouté avant la couche de décision finale.
- **Ajustements** : Le taux d'apprentissage de l'optimiseur a été ajusté pour stabiliser la convergence et accélérer l'apprentissage.

B. Performance

- **Précision** : Le modèle a convergé vers une **précision de validation (val_accuracy) de ~97%**, démontrant une excellente performance pour identifier un instrument seul.
- **Sauvegarde** : Le modèle final a été enregistré sous `best_instrument_classifier.keras` en utilisant la technique du `ModelCheckpoint` pour garantir la sauvegarde des meilleurs poids jamais atteints durant l'entraînement.



III. Détection de Plusieurs Instruments (Classification Multi-Label)

Cette partie répond à la tâche de détection de la combinaison d'instruments jouant ensemble.

A. Création du Dataset Mixé

- Volume de Données :** Pour la complexité de la tâche, un grand dataset de **~26 000 fichiers .wav** a été généré, couvrant toutes les combinaisons de 2 à 5 instruments parmi les 5 cibles (Violon, Piano, Cymbales, Flûte, Vibraphone).
- Mixage Amélioré :** La technique de mixage a été cruciale : chaque piste audio est **normalisée individuellement** avant d'être additionnée, garantissant que les instruments plus faibles (comme la flûte ou le vibraphone) ne soient pas complètement masqués par les sons dominants (Piano, Cymbales).

B. Entraînement et Difficultés

- Architecture :** Pour cette tâche, le modèle a été adapté pour la classification multi-label. La dernière couche a été modifiée pour avoir **5 neurones** (un par instrument cible) avec une activation **sigmoid** (probabilité de présence) et une perte de type binary crossentropy.
- Stratégie :** Nous avons choisi d'entraîner le modèle **depuis zéro** sur les spectrogrammes 3 canaux mixés.
- Résultat peu satisfaisant :** Les tests ont montré que ce modèle échoue à cause du **masquage sonore**. Il s'est concentré uniquement sur les caractéristiques dominantes (Cymbales et Piano) et a ignoré les instruments plus faibles (Flûte, Vibraphone), prouvant que même la richesse des spectrogrammes 3 canaux n'était pas suffisante pour vaincre le bruit.

