Développez une preuve de concept





Contenu de la présentation



02

03

04

05

Choix du jeu de données Choix de modèle Nouveaux concepts et techniques Résultats

Conclusion



Choix du jeu de données





MAESTRO

MAESTRO pour (MIDI and Audio Edited for Synchronous TRacks and Organization), est un ensemble de données publié par Google.

Il contient des enregistrements audio de performances de piano, accompagnés de transcriptions MIDI synchronisées.

Il provient d'enregistrement durant des compétitions de piano International "Piano-e-Compétition" et est souvent utilisé pour des tâches de transcription automatique de musique.

Ses caractéristiques:

- Enregistrements de haute qualité (44.1 kHz en WAV),
- Annotations MIDI alignées avec précision sur l'audio,
- Une large diversité de styles et de tempos, avec plus de 200 heures de musique jouée par des pianistes professionnels.

O

Choix de modèle





Basic Pitch

Basic Pitch est un outil de transcription musicale automatique développé par Spotify. Il utilise un modèle de deep learning basé sur des réseaux neuronaux pour convertir des enregistrements audio en fichiers MIDI. Contrairement à d'autres modèles similaires, Basic Pitch est conçu pour être rapide, léger et robuste, tout en fonctionnant sur une large gamme d'instruments.

Ses caractéristiques:

- Multitimbralité : Peut détecter plusieurs notes jouées simultanément (polyphonie).
- Léger et rapide : Moins gourmand en ressources que d'autres solutions de transcription audio vers MIDI.
- Précision améliorée : Réduction du bruit et meilleure détection des notes par rapport aux modèles classiques.
- Open-source : Disponible sur GitHub et facile à intégrer dans un projet Python.
- Compatibilité large : Peut être utilisé sur du piano, de la guitare, et d'autres instruments mélodiques..



Baseline: Melodia

Melodia est un algorithme de détection de mélodie principale développé par Justin Salamon et ses collègues. Il est conçu pour extraire la hauteur fondamentale (F0) d'une mélodie dominante à partir d'un fichier audio





Sources

 Salamon, J., & Gómez, E. (2012). "Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics." IEEE Transactions on Audio, Speech, and Language Processing

Hawthorne, C., et al. (2018). "Onsets and Frames: Dual-Objective Piano Transcription."
ISMIR 2018.

Bittner, R., Bosch, J. J., Rubinstein, D., Meseguer-Brocal, G., & Ewert, S. (2022). "A
Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and
Multipitch Estimation." ICASSP 2022.

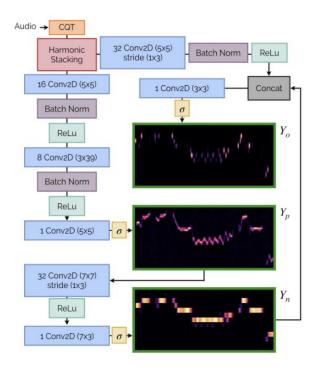


Nouveaux concepts et techniques





CNN



Basic Pitch utilise une architecture de **réseau de neurones convolutifs (CNN)** optimisée pour la transcription musicale. Son pipeline de traitement se divise en plusieurs étapes clés :

- Extraction des caractéristiques audio : Utilisation de spectrogrammes de type Constant-Q Transform (CQT) pour capturer les caractéristiques musicales.
- Modélisation multi-tâches : Le modèle prédit simultanément trois sorties :
 - Présence de notes : Détecte si une note est jouée.
 - Détection d'onsets : Identifie les débuts des notes.
 - Suivi de la hauteur tonale : Suit l'évolution des fréquences.

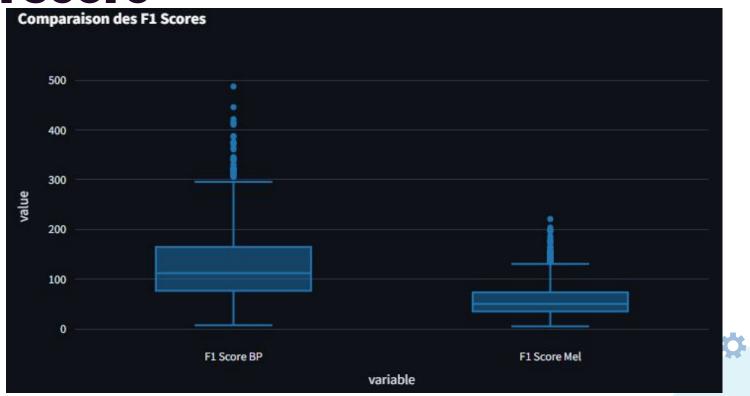
O

Résultats

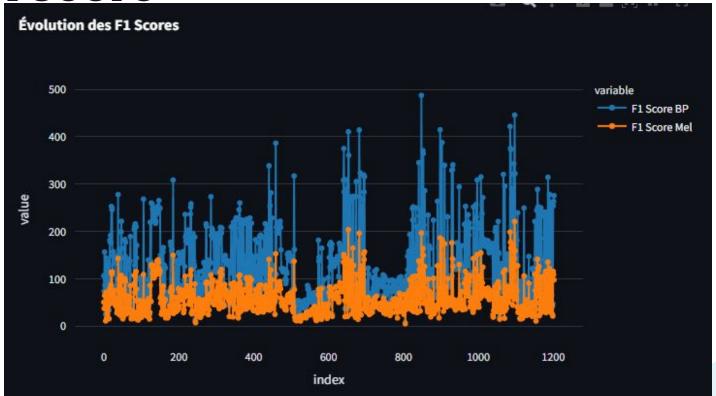




F1 Score



F1 Score





Demo Streamlit





O

Conclusion





Conclusion

Les resultat sont bon au niveaux des metrics mais la ou l'on ressent le plus l'amélioration, c'est lors d'une écoute comparative, comme la musique a une notion d'interprétation qui est difficilement évaluable par des métriques car un point de vue artistique est complexe a automatiser.

Le point d'amélioration , le respect des temps est parfois suffisamment proche pour être juste d'un point de vue metrics, mais extrêmement peux décaler pour que cela choque l'oreil et que cela sonne faux

Un second modèle spécialisé dans le recalage minime des notes dans le temps permettra de changer de quelques microsecondes les notes et permettra un amélioration significative des résultats sans que les metrics soit impacté.