

Musikinstrumentenerkennung mit KI

Können KIs Musikinstrumente erkennen? Dies sollte im Rahmen des Seminars zur *KI-gestützten Audioanalyse von Musik und Soundscapes* an der Hochschule für Musik Franz Liszt Weimar erprobt werden. Ziel war die Entwicklung einer KI, die in der Lage ist, Musikinstrumente zu erkennen. Dazu wurden die Algorithmen des *Maschinellen Lernens*, dem *Deep Learning* und der *Audiosignalverarbeitung* kombiniert.

Die KI basiert auf einem Datensatz mit 30 wav-Dateien pro Musikinstrument, die jeweils eine Länge von 12 Sekunden haben. Die Audiodaten wurden in Form von Mel-Spektrogrammen (Abb.1) dargestellt, die Merkmale des Audiosignals, wie Tonhöhe und Lautstärke, visuell verdeutlichen.

Danach erstellt die KI Klassenlabels. Sie extrahiert Merkmale zu den Gruppen *Violine*, *Viola*, *Cello* und *Klavier*, welche sich in den Überordnern *Streichinstrumente* und *Tastentinstrumente* finden. Natürlich kann die KI um beliebig viele Instrumente erweitert werden, solange der Datensatz entsprechend ausgebaut wird. Daraufhin werden drei verschiedene Datenvektoren generiert: *all_file_id*, *all_class_id* und *all_file_num_in_class*. Die Datenvektoren repräsentieren die Identität der Musikaufnahme (*all_file_id*), die Klasse, zu der sie gehört (*all_class_id*) und die Nummer der Aufnahme innerhalb ihrer Klasse (*all_file_num_in_class*). Am Ende werden diese Datenvektoren zu einem einzigen Datenvektor zusammengefasst, der die Informationen für alle Musikaufnahmen enthält und mit einem Plot visualisiert (Abb.2).

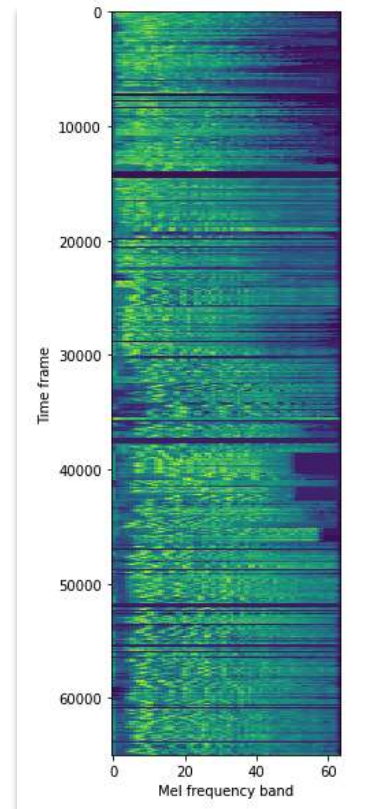


Abbildung 1: Zeigt das Mel-Spektrogramm aller Dateien des Datensatzes mit den Instrumenten Violine, Viola, Cello und Piano

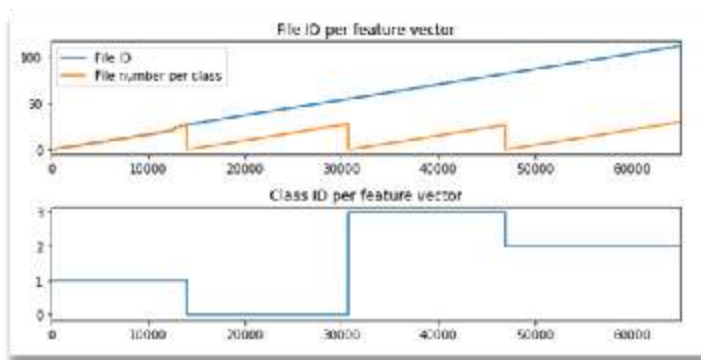


Abbildung 2: Datenvektoren

Die Daten, die für das Training und Testen der KI verwendet wurden, werden nun sorgfältig aufgeteilt. Dabei werden die ersten 20 Dateien der jeweiligen Klassenlabels als Trainingsdaten, und die weiteren 10 Dateien als Testdaten definiert. Anschließend wurden 40 Testdurchläufe durchgeführt, um die Leistungsfähigkeit der KI zu bewerten. Das Ergebnis war ein Accuracy score von 0.7568, was bedeutet, dass die KI in 75% der Fälle erfolgreich in der Lage war, die vier Musikinstrumente korrekt zu identifizieren.

Nun ist die KI bereit, auf für sie bislang unbekannte Audiodateien getestet zu werden. Zur Verdeutlichung werden ihr 3 verschiedene Audiodateien bereitgestellt: a) Audiodatei mit dem solistischen Spielen einer Violine (Abb.3) b) Duett zwischen Cello und Klavier (Abb.4) c) Zusammengeschnittene Audiodatei mit solistischen Stellen der Musikinstrumente in der Reihenfolge Violine, Klavier, Cello, Viola (Abb.5).

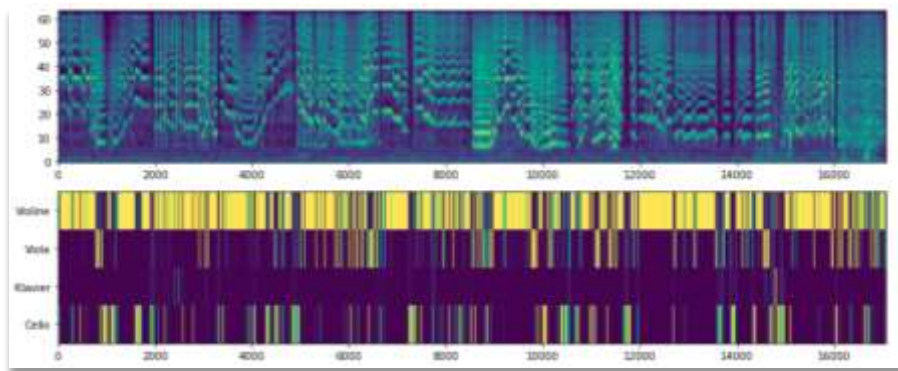


Abbildung 3: Audiobeispiel mit Instrumentensoli Violine

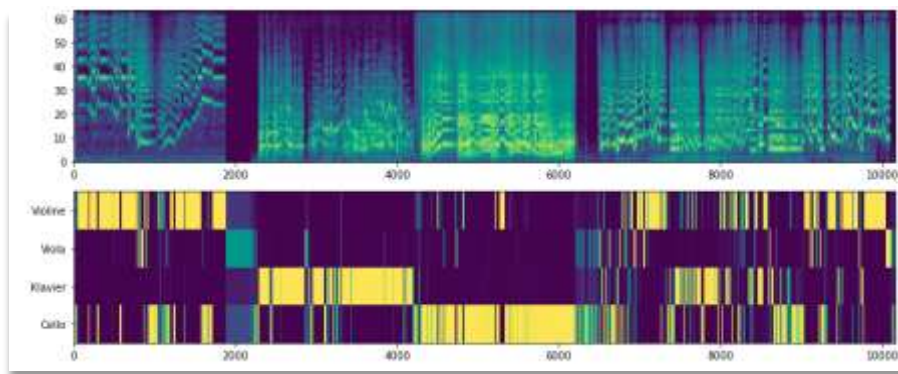
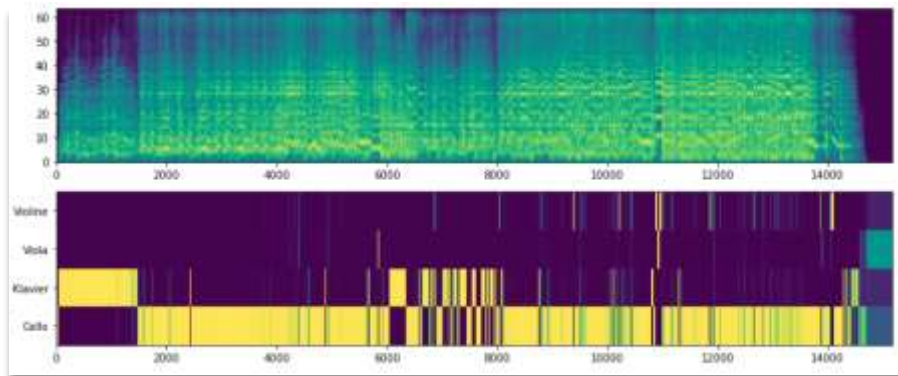


Abbildung 5: Audiobeispiel mit Instrumentensoli in der Reihenfolge Violine, Klavier, Cello, Viola

Die Abbildungen zeigen deutlich, dass die KI die eingepflegten Musikinstrumente überwiegend gut erkennen kann. Probleme treten bei auf, sobald mehrere Instrumente gleichzeitig spielen. Die KI ist auf exekutive Entscheidungen trainiert worden, sodass sie nicht mehrere Instrumente gleichzeitig erkennen kann, die Einrichtung eines „Multilable-Klassifikators“ würde dies vielleicht möglich machen. Auch ist für die KI das Erkennen der Viola schwierig. Durch den stark überschneidenden Frequenzbereich und dem sehr ähnlichen Klang der Streichinstrumente ist dies der KI kaum möglich. Möglicherweise könnte dies durch charakteristischere Dateien zur Spielweise der Viola und einer größeren Menge an Daten im Datensatz verbessert werden. Klar unterscheiden kann die KI zwischen den Streich- und Tasteninstrumenten. Es ist ihr also möglich die Klangfarben zu unterscheiden, was eine gute Basis für weitere Projekte bietet.