

Artificial-Intelligence Based Music Style Transfer

In vergangenen Jahren hat die Künstliche Intelligenz in verschiedenen Forschungsgebieten grosse Fortschritte gemacht. Insbesondere haben generative Modelle starke Resultate in diversen Image Style Transfer Aufgaben gezeigt. In dieser Arbeit werden verschiedene Image-To-Image Modelle für den Audio Style Transfer - beispielsweise von Klassik zu Jazz - verwendet. Zusätzlich werden Massnahmen zur Optimierung der Performanz systematisch untersucht.

In einer einleitenden Analyse passen wir zwei Modelle, CycleGAN und VAEGAN, für den Audio Style Transfer Task an. Die zwei Datensätze die wir dafür benutzen sind ein Instrumentdatensatz, welches Exemplare von Trompete und Violine für Timbre Transfer beinhaltet und ein Musikgenredatensatz für den Transfer von klassischer Musik zu Jazz. Preprocessing beinhaltet die Konvertierung von originalen Audiodateien zu Melspektrogrammen. Die zwei Modelle werden miteinander verglichen und mittels menschlicher Evaluation und Fréchet Audio Distance (FAD) evaluiert. VAEGAN generiert hinreichende Ergebnisse mit beiden Datensätzen, wogegen CycleGAN nur mit dem Instrumentdatensatz arbeiten kann. Daher wird im zweiten Teil der Arbeit nur VAEGAN verwendet.

Im nächsten Teil untersuchen wir die Grenzen der VAEGAN Fähigkeiten bei der Übersetzung von Musikstilen. Die Experimente werden durchgeführt, indem ein Modellparameter isoliert und verschiedene Werte für das Training verwendet werden. Dies zeigt uns die Auswirkung der verschiedenen Parameter und welche Werte die besten Ergebnisse erzielen. Insgesamt würden diese Änderungen an Parametern manchmal zu Verbesserungen des Modells führen, wenn auch nur geringfügig. Die Bewertung der generierten Stilübersetzungen ist eine schwierige Aufgabe und bleibt hauptsächlich subjektiv zu erledigen. Die Parameter mit dem grössten Einfluss waren nmel, welches wichtig für die Konvertierung von Audio zu Melspektrogramm war, der Hyperparameter lambda0 welches den GAN-Loss beeinflusste, die Loss-Funktionen, die Fehler zwischen Bildern berechnete und die für das Training verwendeten Datensätze, die auf einen spezifischeren Jazzstil eingegrenzt werden sollten. Schliesslich wurde der für diese Arbeit verwendete Code auf GitHub veröffentlicht.

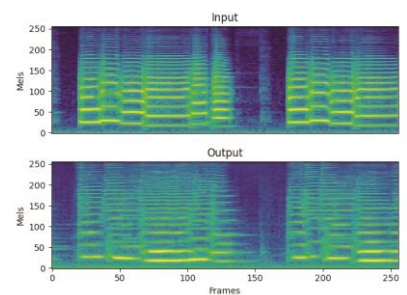


Diplomierende

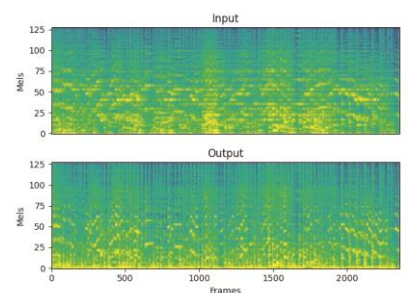
Francis Du
Evan Hirschi

Dozierende

Martin Loeser
Matthias Rosenthal



Audio Style Transfer von Trompete zu Violine mit dem CycleGAN Modell



Audio Style Transfer von Klassik zu Jazz mit dem VAEGAN Modell