

Magister-/Master-/Diplomarbeit

Deconvolution and sonification of features learned by neural networks for musical instrument recognition

Künstliche neuronale Netze (ANNs) haben sich im Bereich des maschinellen Lernens für Audiodaten als erfolgreichstes Werkzeug mit hoher Klassifikationsrate etabliert. Klassischerweise wird das Audiosignal mittels Fouriertransformation zunächst in ein Spektrogramm umgewandelt und dieses schließlich zum Training von Convolutional Neural Networks (CNNs) oder Convolutional Recurrent Neural Networks (CRNNs), deren Ursprung vor allem in der Bilderkennung liegt, genutzt. Ein bedeutender Nachteil von ANNs besteht aus wissenschaftlicher Sicht in der schweren Interpretierbarkeit des von ANNs tatsächlich gelernten Inhalts. Dieses fachübergreifende Problem wurde in der Bildverarbeitung u.a. mithilfe der Dekonvolution gelernter Features behandelt, bei welcher durch einzelne Neuronen gelernte Features auf die Input-Ebene zurückgeführt werden können. So kann auf Bildmuster geschlossen werden, auf welche einzelne Neuronen nach dem Training besonders stark ansprechen und welche im Umkehrschluss ein großes Maß an Information bezüglich des Klassifizierungsproblems beinhalten. Zeiler & Fergus (2014) zeigten, dass durch eine solche Visualisierung von Features vorhandene Mängel der Systemarchitektur aufgedeckt werden können und durch entsprechende Anpassungen der Netzwerke die Performance gesteigert werden kann. Ein Spektrogramm kann zwar technisch als Bild verstanden werden kann, allerdings folgt dessen Inhalt einer anderen Logik: Während bei einem optischen Bild die Elemente eines Objekts für gewöhnlich direkten räumlichen Bezug haben, streckt sich etwa ein Klangmuster für gewöhnlich über mehrere räumlich Obertöne über die Frequenz aus.

Ein alternatives Mittel um dieses Problem anzugehen stellt die Sonifikation, also die Rückumwandlung gelernter Inhalte in Klang dar. Die Umsetzung dieser, welche die logische Konsequenz in Analogie zur Dekonvolution von Bildmaterial darstellt, ist bisher noch sehr wenig erforscht und konnte und es konnte bisher wenig aussagekräftiges Material geschaffen werden. Im Bereich der Music Information Retrieval präsentieren lediglich Choi et al. (2016) eine Dekonvolution und eine folgende Sonifikation gelernter Genre-Features eines CNNs, die Ergebnisse sind hier allerdings noch wenig aussagekräftig. Die Entwicklung eines Tools für die Sonifikation eines Instrumenten-Classifiers soll Inhalt dieser Arbeit sein und klareren Aussagen über den gelernten Inhalt liefern. Der Versuch der Sonifikation verschiedener NN-Architekturen (CNN, RNN, CRNN, TCN) soll untersucht sowie die Effekte der verschiedenen Netzwerk-Parameter getestet werden. Schließlich soll eine quantitative Evaluation der Ergebnisse einer durch die Erkenntnisse der Sonifikation angepassten Systemarchitektur anhand eines aussagekräftigen Datensatzes (z.B. MUSDB2018) soll zusätzlich vorgenommen werden. Die bevorzugte Programmiersprache ist Python.

Literatur

Keunwoo Choi, George Fazekas & Mark Sandler (2016). Explaining Deep Convolutional Neural Networks on Music Classification. arXiv:1607.02444.

Keunwoo Choi, George Fazekas, Mark Sandler, & Jeonghee Kim (2015). Auralisation of deep convolutional neural networks: Listening to learned features. ISMIR latebreaking session.

Matthew D Zeiler & Rob Fergus (2014). Visualizing and understanding convolutional networks. Computer Vision–ECCV 2014, pages 818–833. Springer.

Voraussetzungen

Gute Kenntnisse der Signalverarbeitung und des Maschinellen Lernens (inkl. Neuronale Netzwerke); Python.

Betreuung

Roman Gebhardt, rmngebhardt@gmail.com

Athanasios Lykartsis, EN107, athanasios.lykartsis@tu-berlin.de