

STRESZCZENIE

Określanie dźwięków instrumentów bywa często nazywane kwestią subiektywną, ale istnieją pewne cechy charakterystyczne pozwalające na dywersyfikację ich brzmień. Różnice pomiędzy nimi wywodzą się z właściwości fizycznych tych fal (drgań). Łatwy oraz szybki dostęp do określania tych brzmień mógłby zagwarantować lepszy dobór instrumentów do nastroju i charakterystyki dzieła muzycznego. Praca ma na celu znalezienie tych różnic pomiędzy brzmieniami, oraz przygotowaniu urządzenia które pozwoliłoby na określenie ich rodzaju. Do wykonania projektu wybrano instrument klawiszowy, a następnie zebrano próbki klawisza A4 o częstotliwości 440 Hz dla dwóch powszechnie i umownie określonych brzmień: „jasny”, „miękki”. Po znalezieniu zależności, polegających na intensywnościach oraz liczby składowych drgań harmonicznym dźwięku, przygotowano MEL-spektrogramy danych treningowych. Użyto ich w tworzeniu modelu konwolucyjnej sieci neuronowej zdolnej do rozróżnienia tych klas. Ze względu na ograniczenia czasowe oraz porównaniu modeli tworzonych własnoręcznie, postanowiono skorzystać z frameworku „Teachable Machines” autorstwa Google, udostępnionego do użytku publicznego, pozwalającego na stworzenie modelu będącego w stanie dokonać predykcji z zadowalającą dokładnością. Następnie przygotowano przenośne urządzenie pozwalające na użytkowanie modelu. Wyniki projektu wykazują, że różnice występujące w dźwiękach nie tylko pozwalają na analityczne rozróżnienie barw, ale także na stworzenie modelu będącego w stanie je określać.

Słowa kluczowe:

MEL-spektrogramy, muzyka, analiza dźwięku, sztuczna inteligencja, konwolucyjne sieci neuronowe, programowanie, Python, komputer jednopłytkowy, RaspberryPi, urządzenia embedded.

ABSTRACT

Defining the sounds of instruments is often called a subjective matter, but there are certain characteristics that allow for the diversification of their sounds. The differences between them are derived from the physical properties of these waves (vibrations). Easy and quick access to defining these sounds could guarantee a better selection of instruments for the mood and characteristics of a musical work. The work aims to find these differences between sounds, and to prepare a device that would allow for determining their type. A keyboard instrument was selected for the project, and then samples of the A4 key with a frequency of 440 Hz were collected for two commonly and conventionally defined sounds: "bright", "soft". After finding the relationships, consisting in the intensities and the number of components of harmonic vibrations of the sound, MEL-spectrograms of training data were prepared. They were used to create a convolutional neural network model capable of distinguishing these classes. Due to time constraints and the comparison of models created by hand, it was decided to use the Google "Teachable Machines" framework, made available for public use, which allows the creation of a model capable of making predictions with satisfactory accuracy. Then, a portable device was prepared to allow the use of the model. The results of the project show that the differences occurring in sounds not only allow for analytical distinction of colors, but also for the creation of a model capable of defining them.

Keywords:

MEL-spectrograms, music, sound analysis, artificial intelligence, convolutional neural networks, programming, Python, single-board computer, RaspberryPi, embedded devices.

SPIS TREŚCI

Wykaz ważniejszych słów i skrótów	6
1. Cel pracy	7
2. Wstęp teoretyczny	8
2.1. Dźwięk oraz jego analiza	9
2.1.1. <i>Charakterystyka dźwięku oraz badanego instrumentu</i>	9
2.1.2. <i>MEL-spektrogramy</i>	10
2.1.3. <i>Transformata Fouriera</i>	11
2.1.4. <i>Analiza dźwięku</i>	12
2.2. Sztuczne sieci neuronowe	13
2.2.1. <i>Działanie oraz zastosowanie sztucznych sieci neuronowych</i>	13
2.2.2. <i>Konwolucyjne sieci neuronowe i ich zastosowanie w projekcie</i>	14
2.2.3. <i>Framework Teachable machines</i>	15
2.3. Język programowania Python	16
2.3.1. <i>Charakterystyka języka</i>	16
2.3.2. <i>Wykorzystane biblioteki</i>	17
2.4. Urządzenia embedded	18
2.4.1. <i>Ogólna charakterystyka</i>	18
2.4.2. <i>RaspberryPi</i>	19
3. Przebieg badania i metodologia	20
3.1. Dane	21
3.1.1. <i>Pozyskiwanie danych</i>	22
3.1.2. <i>Analiza</i>	23
3.1.3. <i>Przygotowanie zbiorów treningowych</i>	24
3.2. Stworzenie modelu sieci neuronowej	25
3.3. Aplikacja do klasyfikacji dźwięku	26
3.3.1. <i>Struktura i działanie aplikacji</i>	27
3.3.2. <i>Elementy graficzne</i>	28
3.3.3. <i>Implementacja i funkcjonalność modelu modelu sieci</i>	29
3.4. Obudowa urządzenia	30
4. Podsumowanie	31
Wykaz literatury	32
Wykaz rysunków	33
Wykaz tabel	34
Materiały dodatkowe: Repozytorium projektu	35

WYKAZ WAŻNIEJSZYCH SŁÓW I SKRÓTÓW

CNN – konwolucyjna sieć neuronowa (od ang. Convolutional Neural Network)

zmienna – w języku programowania oznacza to nośnik danych w programie bądź funkcji

funkcja – w języku programowania jest to definiowalna metoda, która przyjmuje zmienne oraz wykonuje operacje logiczne

epoka – jedno kompletne przejście zestawu danych treningowych przez algorytm

barwa dźwięku – inne określenie na charakterystykę brzmienia instrumentu

1. CEL PRACY

2. WSTĘP TEORETYCZNY

2.1. Dźwięk oraz jego analiza

2.1.1. *Charakterystyka dźwięku oraz badanego instrumentu*

2.1.2. MEL-spektrogramy

2.1.3. Transformata Fouriera

2.1.4. Analiza dźwięku

2.2. Sztuczne sieci neuronowe

2.2.1. Działanie oraz zastosowanie sztucznych sieci neuronowych

2.2.2. Konwolucyjne sieci neuronowe i ich zastosowanie w projekcie

2.2.3. Framework Teachable machines

2.3. Język programowania Python

2.3.1. *Charakterystyka języka*

2.3.2. Wykorzystane biblioteki

2.4. Urządzenia embedded

2.4.1. *Ogólna charakterystyka*

2.4.2. RaspberryPi

3. PRZEBIEG BADANIA I METODOLOGIA

Gdyby podczas badań bądź użytku pojawiła się inna częstotliwość niż badana postanowiono dodać trzecią kategorię brzmienia pod nazwą „niekreślone”. Po zebraniu danych dźwiękowych zamieniono je na dane liczbowe używając transformaty Fouriera, a następnie sporządzono z nich MEL-spektrogramy. Po zauważeniu różnic w występujących częstotliwościach harmonicznym oraz ich intensywnościach, zebrano próbki dźwiękowe z pianina cyfrowego dla każdego z brzmień, przygotowano model konwolucyjnej sieci neuronowej a następnie wytrenowano go przy użyciu MEL-spektrogramów, w ilości 240 plików dla każdego z przypadków. Dla przygotowanego modelu w języku Python została napisana aplikacja analizująca dźwięk instrumentu klawiszowego, informująca o predykcji klasy oraz pewności modelu podana w procentach. Została również zaimplementowana graficzna pomoc w obsłudze oraz wyświetlanie MEL-spektrogramu próbki. Jako urządzenie obliczeniowe, które pozwoliłoby na budowę przenośnego urządzenia o możliwości wyświetlania graficznych funkcjonalności aplikacji, wykorzystano jednopłytkowy komputer RaspberryPi model 4B.

3.1. Dane

3.1.1. Pozyskiwanie danych

3.1.2. Analiza

3.1.3. Przygotowanie zbiorów treningowych

3.2. Stworzenie modelu sieci neuronowej

3.3. Aplikacja do klasyfikacji dźwięku

3.3.1. Struktura i działanie aplikacji

3.3.2. *Elementy graficzne*

3.3.3. Implementacja i funkcjonalność modelu modelu sieci

3.4. Obudowa urządzenia

4. PODSUMOWANIE

WYKAZ LITERATURY

WYKAZ RYSUNKÓW

WYKAZ TABEL

Materiały dodatkowe: Repozytorium projektu

<https://github.com/Kiryl24/Master-thesis-project>