

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

**AGH UNIVERSITY OF SCIENCE
AND TECHNOLOGY**

AGH

Wykorzystanie uczenia maszynowego do identyfikacji metody kompresji sygnału akustycznego

Autor: Szymon Mikulicz

Opiekun: dr hab. inż. Bartłomiej Borkowski

Recenzent: prof. dr hab. inż. Jerzy Wiciak

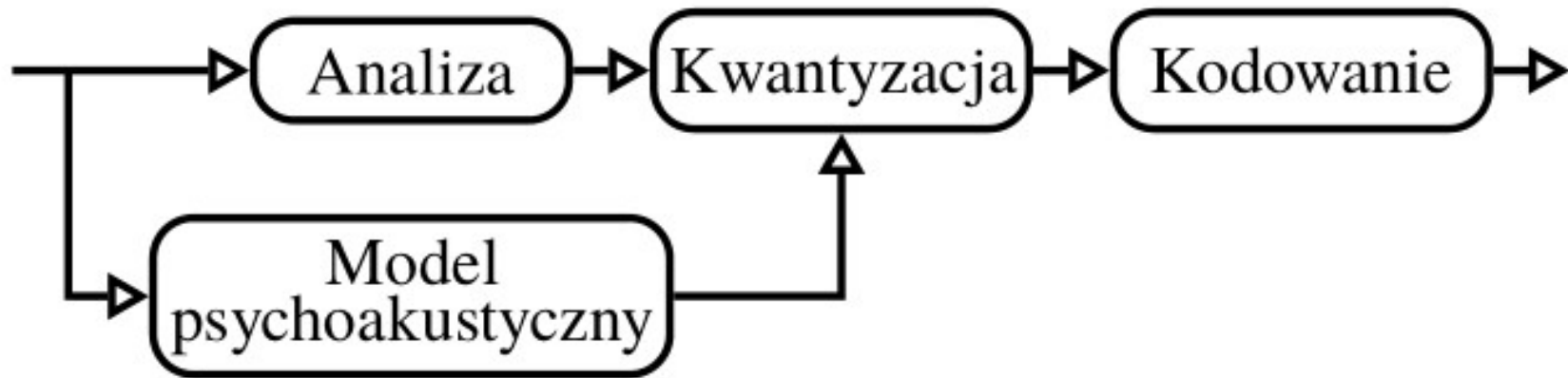
Data: 5.09.2019 r.

Wprowadzenie

- » Kompresja audio stosowana jest w celu zmniejszenia rozmiaru pliku
- » Kompresja może być stratna lub bezstratna
- » W sprzedaży mogą występować pliki bezstratne uzyskane przez dekompresję stratnych



Kompresja stratna audio (1)



- » Proces jest nieodwracalny
- » Skutkuje pogorszeniem jakości sygnału
- » Poziom kompresji może być regulowany przez użytkownika

Obecne rozwiązania

- » Narzędzia: Lossless Audio Checker, True Audio Checker
- » Algorytmicznie wykrywają artefakty powstałe w kompresji MPEG (MP3, AAC)
- » Prowadzone są badania nad uniwersalnymi algorytmami (2) oraz uczeniem maszynowym (3)

Konwolucyjne Sieci Neuronowe

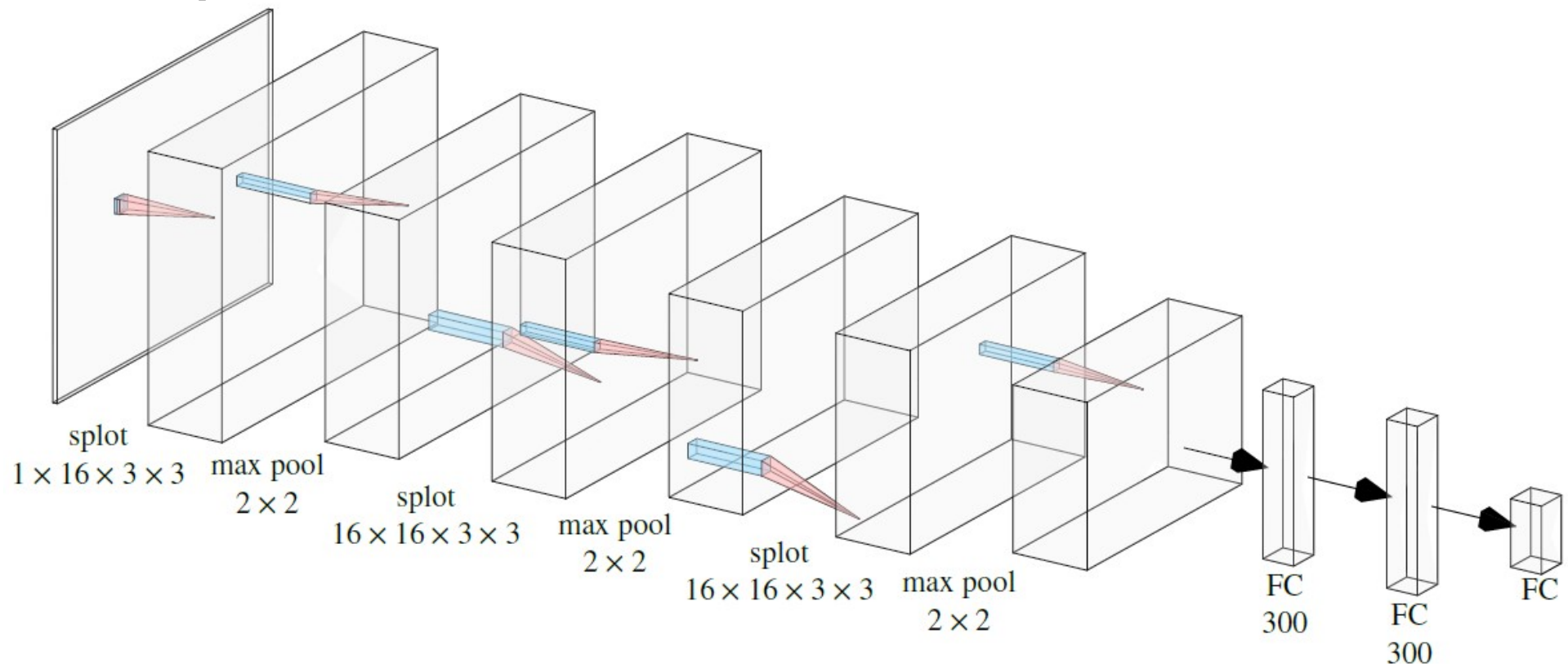
- » Wykorzystywane do rozpoznawania obrazów
- » Składają się z warstw splatających i warstw łączących (pooling)

$$\left\langle \begin{bmatrix} 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 3 \\ 3 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \mathbf{J} \right\rangle_F = \begin{bmatrix} 12 \end{bmatrix} \quad \left\langle \begin{bmatrix} 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 3 \\ 3 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \mathbf{J} \right\rangle_F = \begin{bmatrix} 12 & 12 \end{bmatrix}$$

$$\max \left\{ \begin{bmatrix} 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 3 \\ 3 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix} \right\} = \begin{bmatrix} 3 \end{bmatrix} \quad \max \left\{ \begin{bmatrix} 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 3 \\ 3 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix} \right\} = \begin{bmatrix} 3 & 3 \end{bmatrix}$$

Architektura sieci

- » Funkcja aktywacji: ReLU
- » Dropout: 30%, 20%, 20%



Przygotowane narzędzia

- » Skrypt do konstrukcji bazy danych z plików audio
- » Biblioteka do weryfikacji bezstratności fragmentów przed treningiem (2)
- » Skrypt dobierający i przetwarzający dane, prowadzący trening, testujący i przetwarzający wyniki



Dane Treningowe

- » Baza 65536 10-ciosekundowych fragmentów
- » Stopień kompresji regulowany parametrem bitrate
- » 4096 niepoddanych kompresji, po 4096 dla każdej z par formatów MP3, AAC, AC-3, Vorbis, WMA z bitrate'ami 320 kbps, 192 kbps, 128 kbps

Trening

- » 10 epok
- » Wybierano sieć o największej dokładności
- » Optymizator: ADAM
- » Funkcja strat: NLL (Negative Log Likelihood)
- » 3 konfiguracje

Wyniki treningu I: 97,4%

		Format wykryty przez model	
		WAV	Inny
Rzeczywisty format	WAV	—	97,6
	MP3	320	11,6
		192	7,0
		128	3,5
	AAC	320	4,7
		192	0,0
		128	0,0
	Vorbis	320	1,2
		192	0,0
		128	0,0
	WMA	320	12,9
		192	0,0
		128	0,0
	AC-3	320	0,0
		192	0,0
		128	1,2

Wyniki treningu II: 96,1%

		Format wykryty przez model						
		WAV	MP3	AAC	Vorbis	WMA	AC-3	
Rzeczywisty format	WAV	—	94,4	4,9	0,0	0,0	0,5	0,2
	MP3	320	4,9	90,2	0,0	0,0	4,9	0,0
		192	3,5	95,1	0,0	0,0	0,7	0,7
		128	4,9	91,5	2,8	0,0	0,7	0,0
	AAC	320	1,4	2,1	95,1	0,7	0,7	0,0
		192	0,0	0,7	94,4	4,9	0,0	0,0
		128	0,0	0,0	91,5	7,7	0,7	0,0
	Vorbis	320	0,0	0,0	2,8	96,5	0,7	0,0
		192	0,0	0,0	0,0	100,0	0,0	0,0
		128	0,0	0,0	2,1	97,9	0,0	0,0
	WMA	320	2,1	0,0	0,0	0,7	97,2	0,0
		192	0,0	0,0	0,0	0,0	100,0	0,0
		128	0,0	0,0	0,0	0,0	100,0	0,0
	AC-3	320	0,7	0,0	0,0	0,0	0,7	98,6
		192	0,0	0,0	0,0	0,0	1,4	98,6
128		0,0	0,0	0,0	0,0	0,7	99,3	



Wyniki treningu III: 92,4%

Format wykryty przez model

		WAV	MP3			AAC			Vorbis			WMA			AC-3		
		—	320	192	128	320	192	128	320	192	128	320	192	128	320	192	128
WAV	—	96,1	0,4	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,0	0,0	0,0	0,0	0,0	0,0
MP3	320	7,8	89,5	2,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,8	0,0	0,0	0,0	0,0	0,0
	192	4,3	0,0	94,5	0,8	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,0	0,0	0,0	0,0
	128	4,3	0,0	0,4	94,9	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
AAC	320	1,2	0,0	0,0	1,2	94,9	0,8	0,4	0,0	0,0	0,0	1,6	0,0	0,0	0,0	0,0	0,0
	192	0,8	0,0	0,0	0,0	2,3	93,0	1,2	0,0	1,2	0,8	0,8	0,0	0,0	0,0	0,0	0,0
	128	0,0	0,0	0,0	0,0	0,8	1,2	85,5	9,4	2,0	1,2	0,0	0,0	0,0	0,0	0,0	0,0
Vorbis	320	0,0	0,0	0,0	0,0	0,4	1,2	4,7	80,1	10,2	2,0	0,8	0,4	0,0	0,4	0,0	0,0
	192	0,4	0,0	0,0	0,0	0,0	0,4	0,0	5,5	69,9	23,0	0,4	0,4	0,0	0,0	0,0	0,0
	128	0,4	0,0	0,0	0,0	0,4	0,0	0,0	0,0	14,1	84,8	0,0	0,4	0,0	0,0	0,0	0,0
WMA	320	0,4	0,8	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	98,4	0,0	0,0	0,0	0,0	0,0
	192	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,8	99,2	0,0	0,0	0,0	0,0
	128	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	99,6	0,0	0,0	0,0
AC-3	320	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,0	99,2	0,0	0,0
	192	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
	128	0,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,0	0,0	0,0	98,8

Rzeczywisty format

Podsumowanie

- » Sieć rozpoznaje z dużą dokładnością obecnie stosowane metody stratnej kompresji audio
- » Rozwiązanie jest możliwe do zastosowania przez serwisy prowadzące sprzedaż plików audio

Dalsze pracy

- » Większa złożoność sieci
- » Różne enkodery
- » Więcej metod kompresji
- » Inne typy sieci

Wykorzystanie uczenia maszynowego do identyfikacji metody kompresji sygnału akustycznego

<https://github.com/Ashymad/IOLA>

Dziękuję za uwagę

Bibliografia:

- (1) Marina Bosi i Richard E. Goldberg: *Introduction to Digital Audio Coding and Standards*, Norwell, MA, USA: Kluwer Academic Publishers, 2002, isbn: 1402073577
- (2) Bongjun Kim: „Lossy Audio Compression Identification”, w: *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, s. 2459-2463
- (3) Romain Hennequin, Jimena Royo-Letelier i Manuel Moussallam: „Codec independent lossy audio compression detection”, w: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, s. 726–730