

Zenei stílusok felismerése. neurális hálózattal

Önálló labor 2
Készítette: Kenesei Benjámin
Konzulens: Dr. Bank Balázs

Önálló labor célja

Egy olyan program készítése, amely képes zeneszámok műfajainak a felismerésére egy CNN segítségével.



Lépések

Adatbázis keresés

Tanító adatok

Spektogram



Spektogram
finomítás

FFT paraméterek

Melspektogram

Colormap

Frekvencia tartomány



CNN

5 konvolúciós réteg

1 dropout réteg

1 pooling réteg

1 lineáris réteg



Eredmények

31 epoch után

97%-os validációs pontosság

Tanító adatok

9 műfaj: Blues, Classic, Country, Disco, Hiphop, Metal, Pop, Reggae, Rock

GTZAN adatbázis

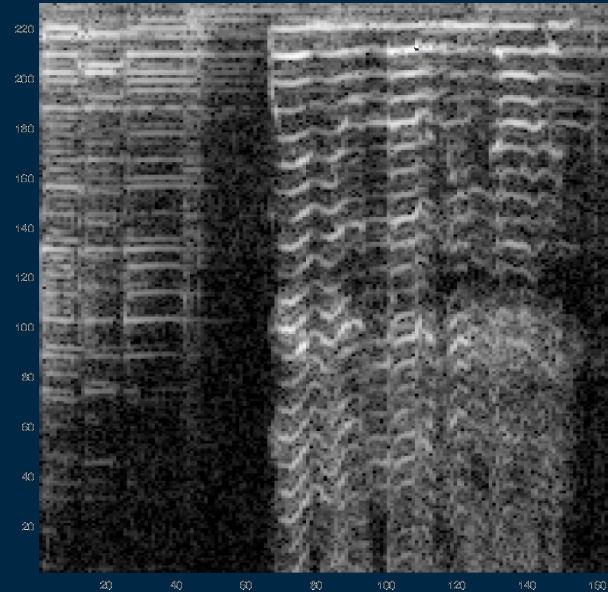
Műfajonként 100 db 30 mp-es zeneszám. -> 1000 db 3mp-es fájl

[\(Music Genre Recognition using Convolutional Neural Networks – Kunal Vaidya\)](#)

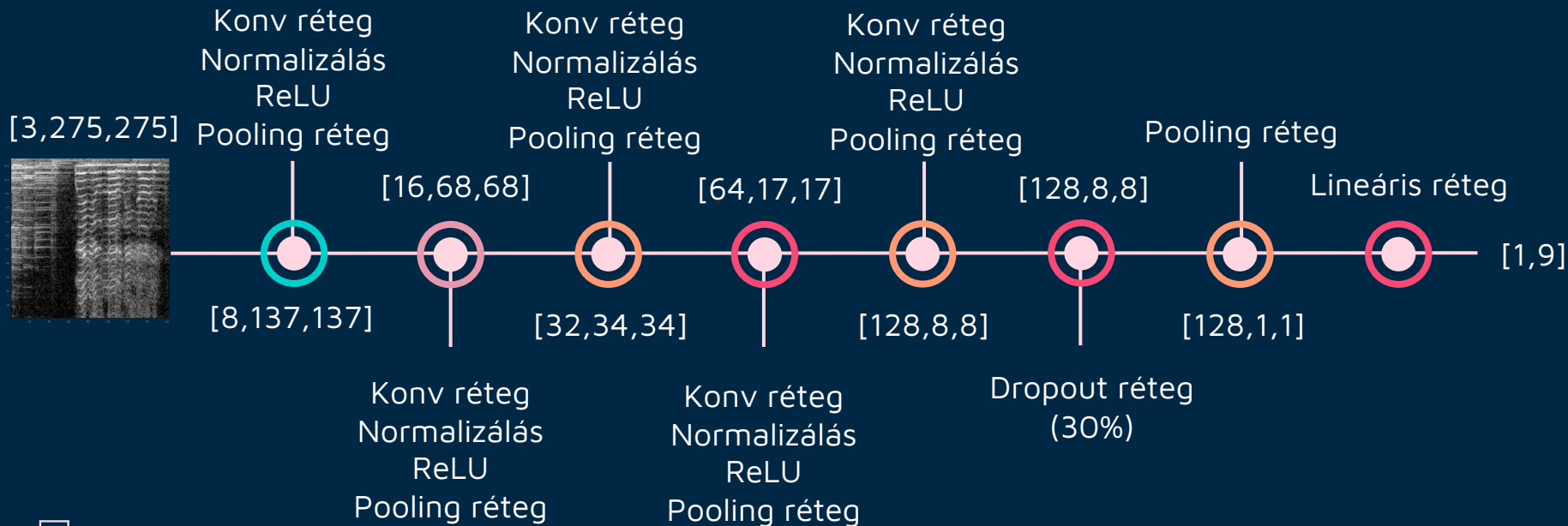
Youtube adatbázis

Hosszú videók feldarabolása -> hasonló mennyiségű tanító adat

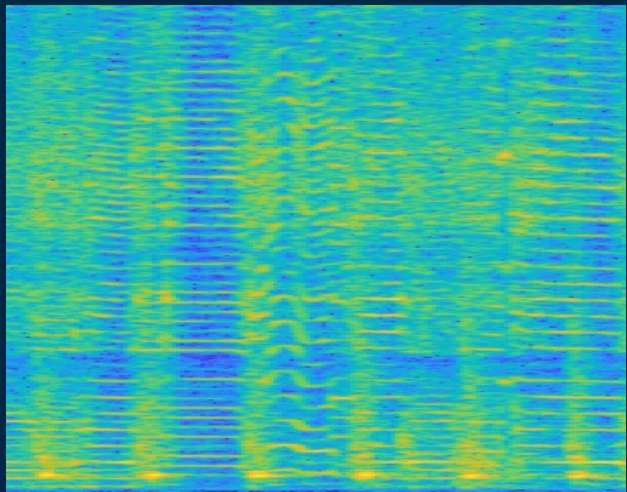
Spektrogram



CNN felépítése

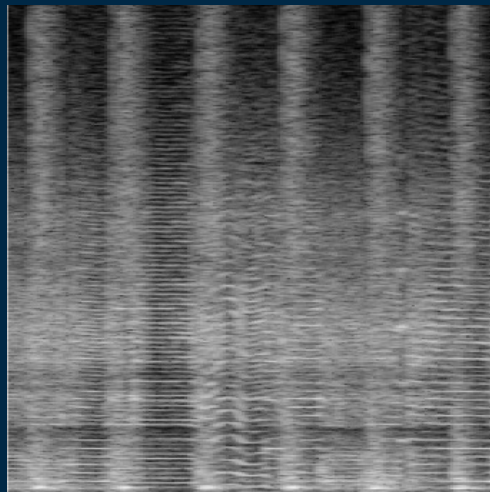


Spektrogram finomítás

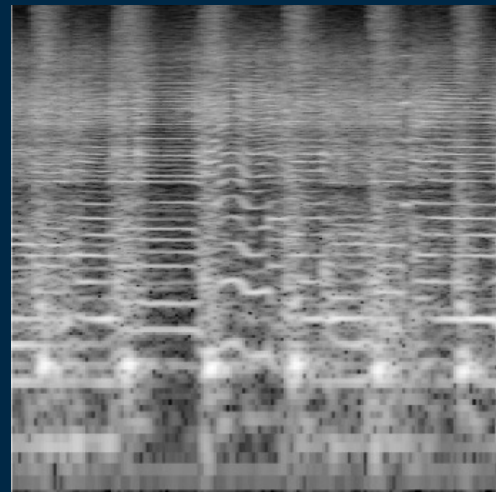


01

Lineáris ábrázolás
0-3000 Hz
Colormap
82%-os pontosság



(lineáris)

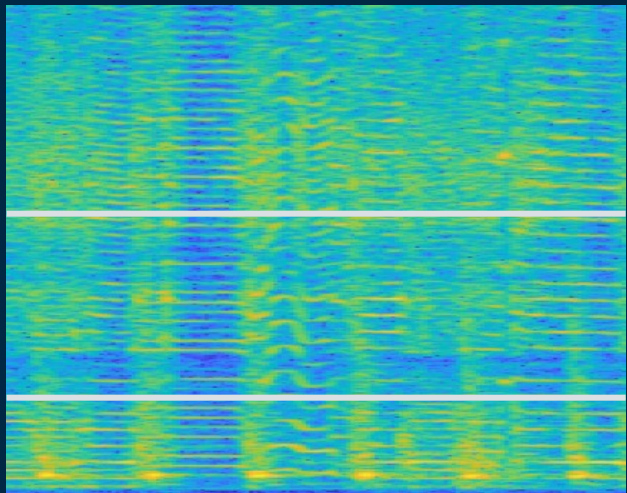


(logaritmus)

02

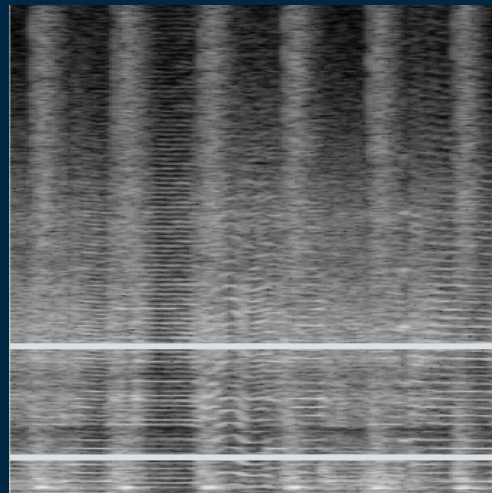
Logaritmus ábrázolás
50-10000 Hz
Grayscale
93%-os pontosság

Spektrogram finomítás



01

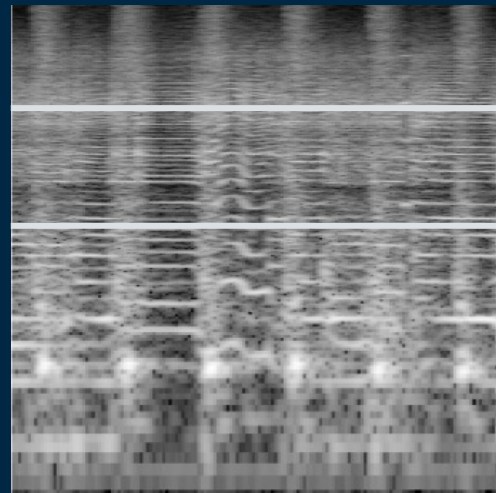
Lineáris ábrázolás
0-3000 Hz
Colormap
82%-os pontosság



(lineáris)

02

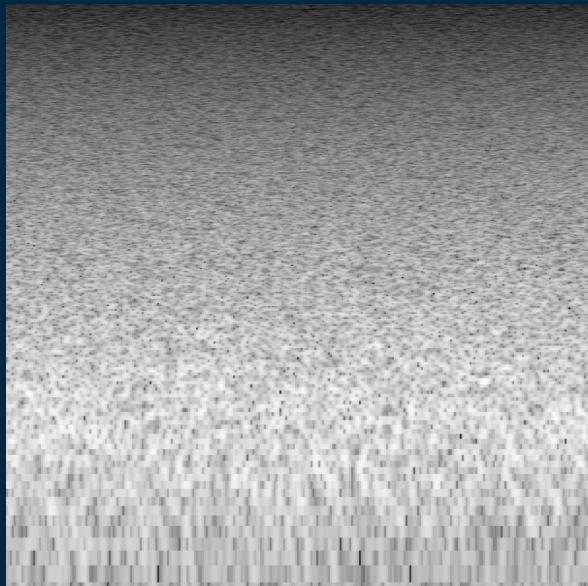
Logaritmusikus ábrázolás
50-10000 Hz
Grayscale
93%-os pontosság



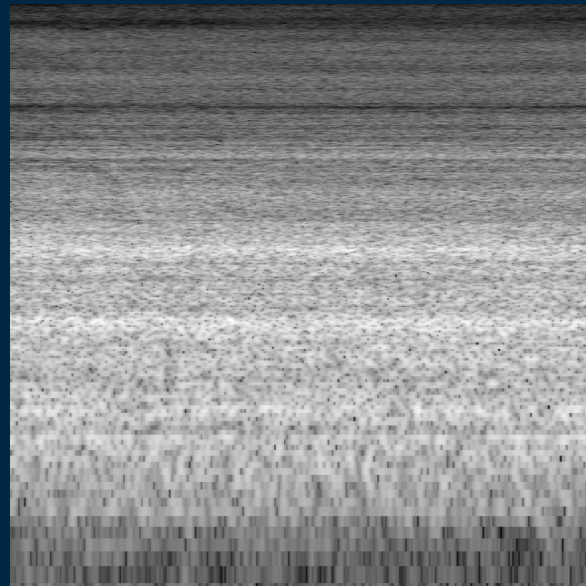
(logaritmusikus)

Rossz minőségű felvételek felismerése

Eredeti fehér zaj



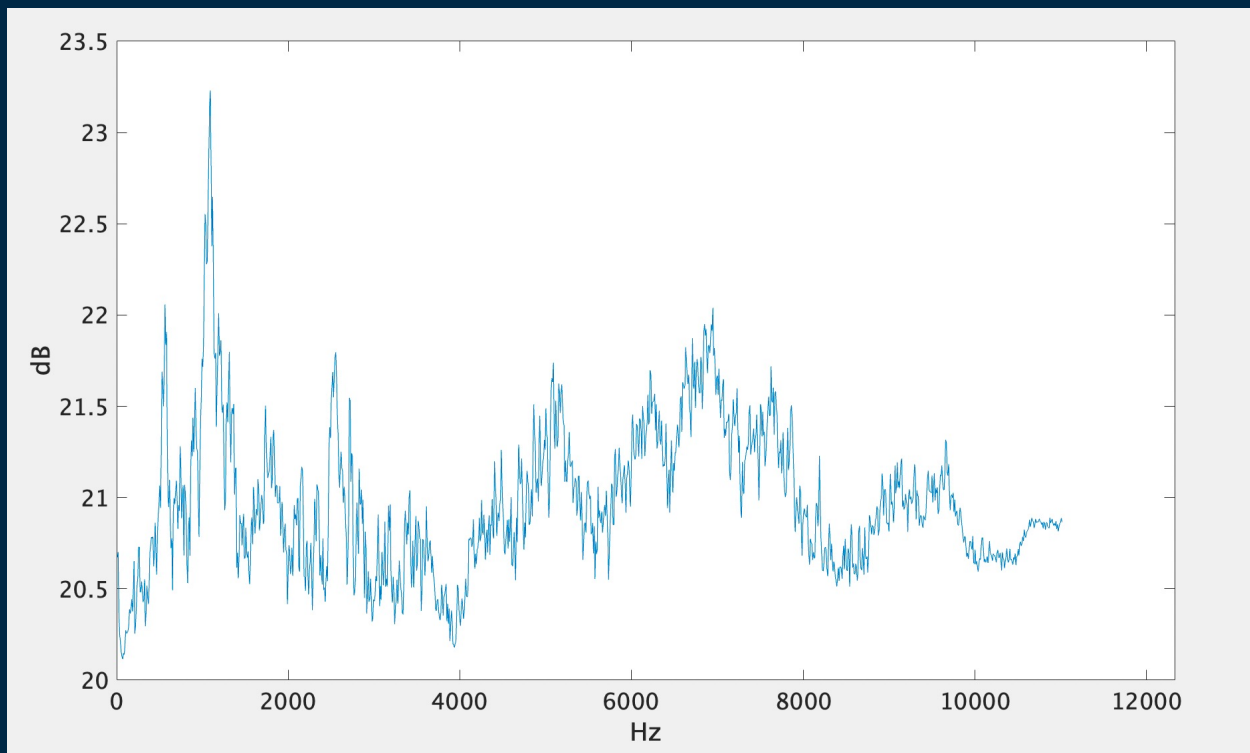
Laptop hangszóról felvett fehér zaj



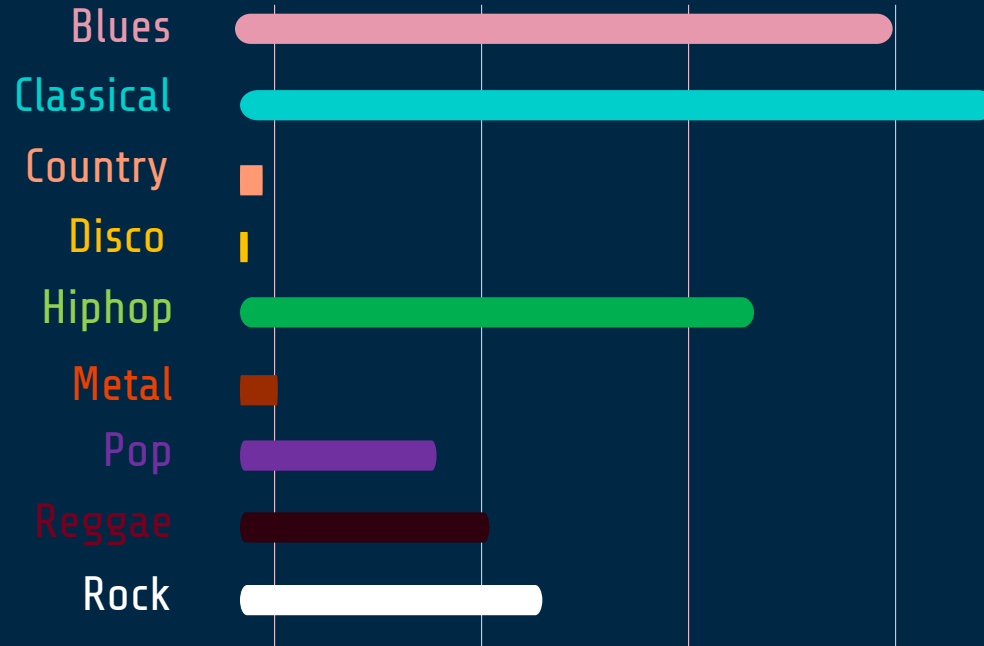
Összesen pár százalékot ront a pontosságon,
ezért nem foglalkoztam vele bővebben

Átviteli függvény számítás

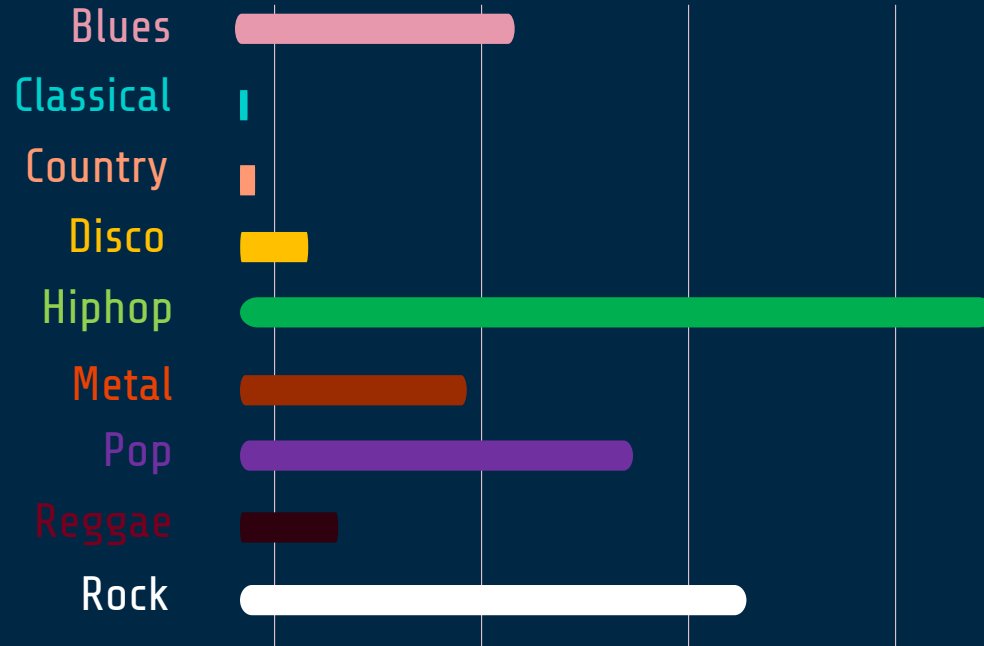
Rossz minőségű felvételek felismerése miatt



Edvard Grieg – In the Hall of the Mountain King



Beastie Boys – No Sleep Till Brooklyn



Confusion

Mátrix

GTZAN adatbázis,
3 mp-es spektrogrammok
~92%-os pontosság

Címke

Predikció

	Blues	Classic	Country	Disco	Hiphop	Metal	Pop	Reggae	Rock
Blues	93.5								3
Classic		99							
Country			94.5					2	1.5
Disco			3.5	86	2		2	2	4
Hiphop					95.5		3		
Metal						94.5			3.5
Pop			2		1.5		90	2.5	2.5
Reggae	1.5		2		1.5		2	91.5	
Rock	2		4	1.5	1.5		2.5	1.5	85.5

Confusion

Mátrix

GTZAN adatbázis,
6 mp-es spektrogrammok
~97%-os pontosság

Címke

Predikció

	Blues	Classic	Country	Disco	Hiphop	Metal	Pop	Reggae	Rock
Blues	98		1		1				
Classic		100							
Country	1		95				3		1
Disco		1		98				1	
Hiphop				2	94		2		2
Metal	1				1	97			1
Pop				1			99		
Reggae							1	99	
Rock	1		1	1				1	96

Pontosság 1., 2., 3. tipp esetén

GTZAN adatbázison tanított, Youtube adatbázison tesztelt háló

	Blues	Classic	Country	Disco	Hiphop	Metal	Pop	Reggae	Rock	AVG*
1	54	85	27	58	69	86	53	65	26	58,1
1,2	70	95	39	83	91	94	66	79	76	77,0
1,2,3	75	97	48	96	97	96	73	86	89	84,1

*A véletlen találgatás esélye 11% lenne

Pontosság 1., 2., 3. tipp esetén

GTZAN + Youtube1 adatbázison tanított,
Youtube2 adatbázison tesztelt háló

	Blues	Classic	Country	Disco	Hiphop	Metal	Pop	Reggae	Rock	AVG*
1	72	93	60	62	58	52	67	74	64	66,9
1,2	87	96	73	80	71	75	83	85	85	81,7
1,2,3	94	98	82	90	88	87	94	91	92	90,7

*A véletlen találgatás esélye 11% lenne

Fejlesztési opciók

- Felhasználóbarát alkalmazás
- Új zenei jellemzők bevonása(spectral rollof, zero crossing rate...)
- Konzisztens adatbázis