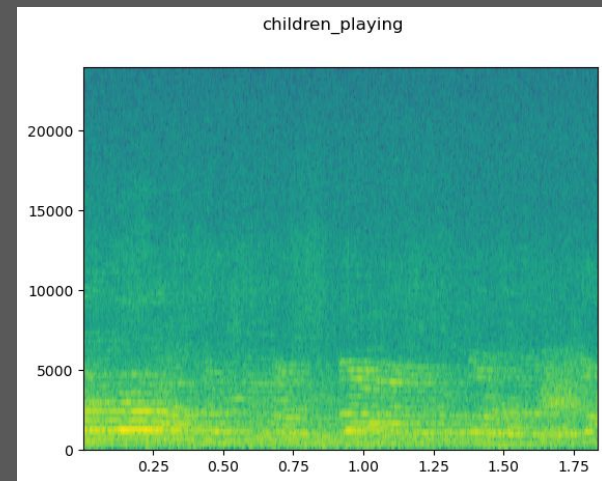
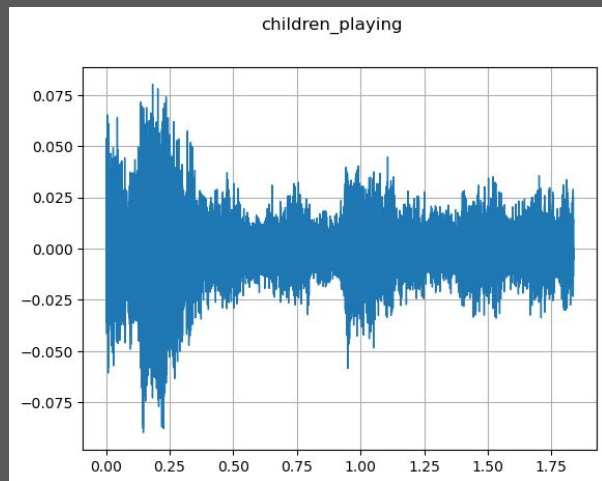
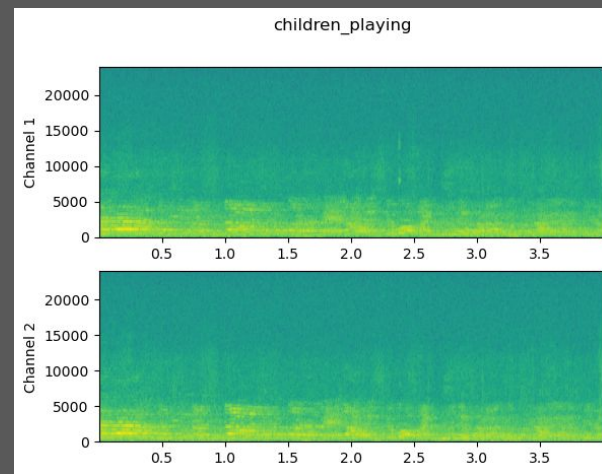
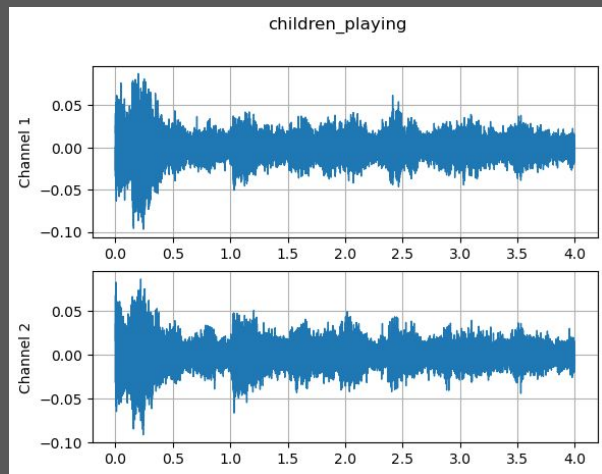


Klasifikacija Urbanih zvukova

Branko Grbić
Željko Milovanović



O podacima

Urban Sound 8K je skup audio podataka koji sadrži 8732 označenih uzoraka (≤ 4 s) 10 klasa urbanih zvukova: air_conditioner, car_horn, children_playing, dog_bark, drilling, engine_idling, gun_shot, jackhammer, siren, and street_music.

Klase su skoro ravnomerno raspoređene:

	index	jackhammer	dog_bark	children_playing	street_music	air_conditioner	drilling	engine_idling	siren	car_horn	gun_shot
0	fold1	120	100	100	100	100	100	96	86	36	35
1	fold2	120	100	100	100	100	100	100	91	42	35
2	fold3	120	100	100	100	100	100	107	119	43	36
3	fold4	120	100	100	100	100	100	107	166	59	38
4	fold5	120	100	100	100	100	100	107	71	98	40
5	fold6	68	100	100	100	100	100	107	74	28	46
6	fold7	76	100	100	100	100	100	106	77	28	51
7	fold8	78	100	100	100	100	100	88	80	30	30
8	fold9	82	100	100	100	100	100	89	82	32	31
9	fold10	96	100	100	100	100	100	93	83	33	32

dog_bark	0.114521
children_playing	0.114521
air_conditioner	0.114521
street_music	0.114521
engine_idling	0.114521
jackhammer	0.114521
drilling	0.114521
siren	0.106390
car_horn	0.049130
gun_shot	0.042831

Motivacija

Cilj ovog projekta je pokazati razliku između ne pretrenirane i pretrenirane Konvolutivne neuronske mreže za klasifikaciju audio materijala.

Cilj je takođe pokazati zašto su pretrenirani modeli dominantniji u industriji.

Pošto podaci nisu homogeni ni na koji način, različit broj audio kanala, drugačija stopa odabiranja, drugačijeg trajanja itd.

Morali smo da se prilagodimo i prevazišemo ove prepreke modifikujući naš skup podataka, ne bi li ga napravili homogenim i adekvatnim ulazom za naše modele.

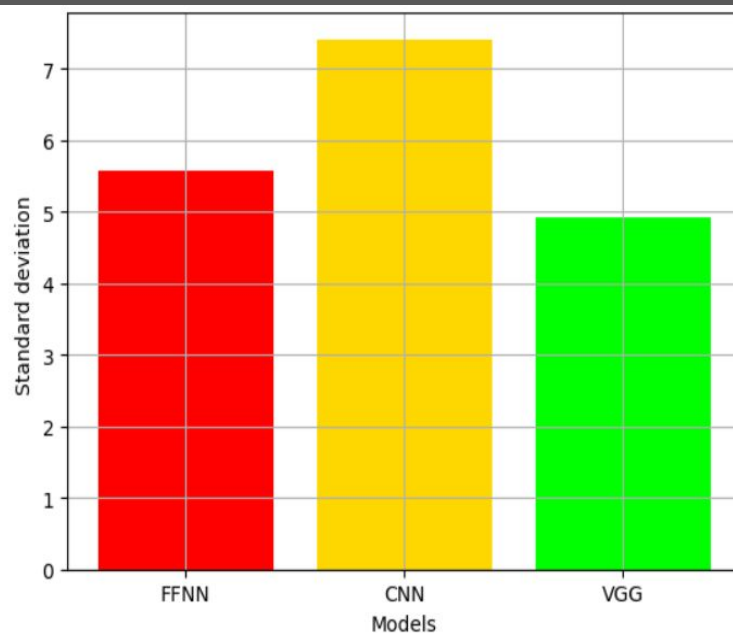
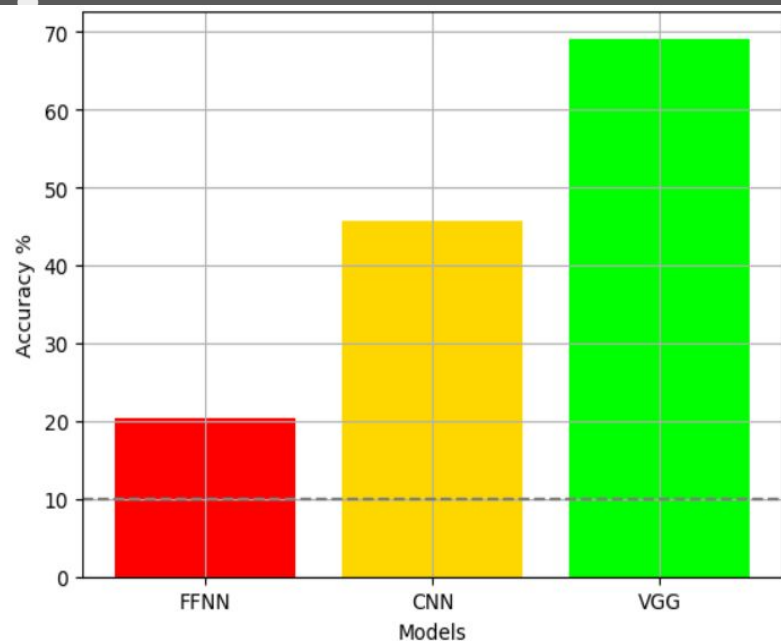
Predprocesiranje podataka

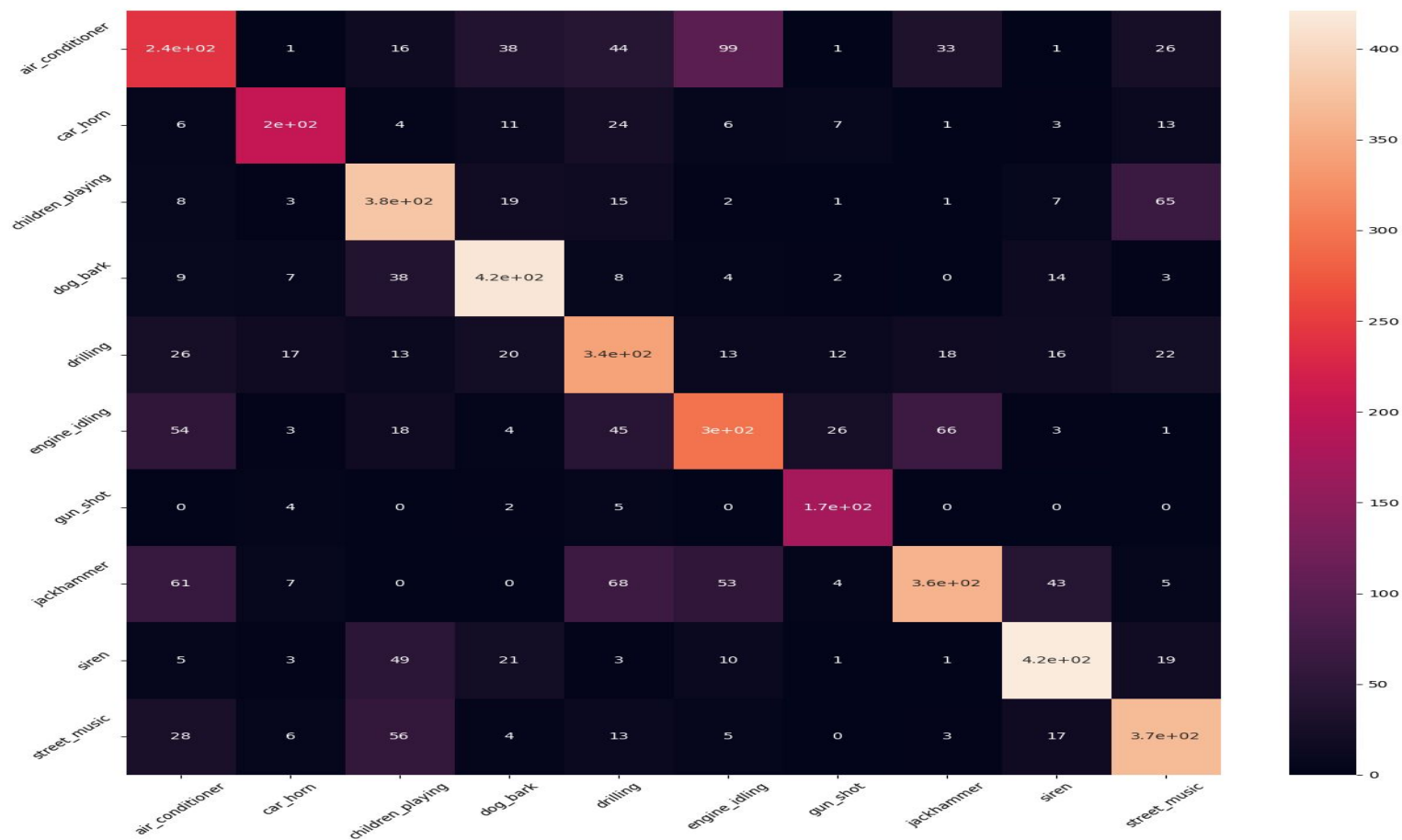
- Ponovno uzorkovanje (Resampling)
- Konvertovanje zvuka u mono
- Skraćivanje
- Dopunjavanje sa desne strane
- Izvlačenje MFCC karakteristika
- Konvertovanje u 3 kanala
- Konvertovanje u tenzor
- 9 foldova za trening, 1 za validaciju

Modeli

- Potpuno povezana neuronska mreža (FFNN)
 - Konvolutivna neuronska mreža sa VGG - olikom arhitekturom (CNN)
 - Pre-trenirani VGG (ImageNet)
-
- Funkcija gubitka - Kategorička unakrsna entropija
 - Optimizator - Adam
 - Stopa učenja - $1e-4$
 - Korak stope učenja - 5 (gamma 0.1)
 - Batch size - 64
 - Broj epoha 9 (VGG, CNN), 13 (FFNN)

Tačnost i standardno odstupanje modela

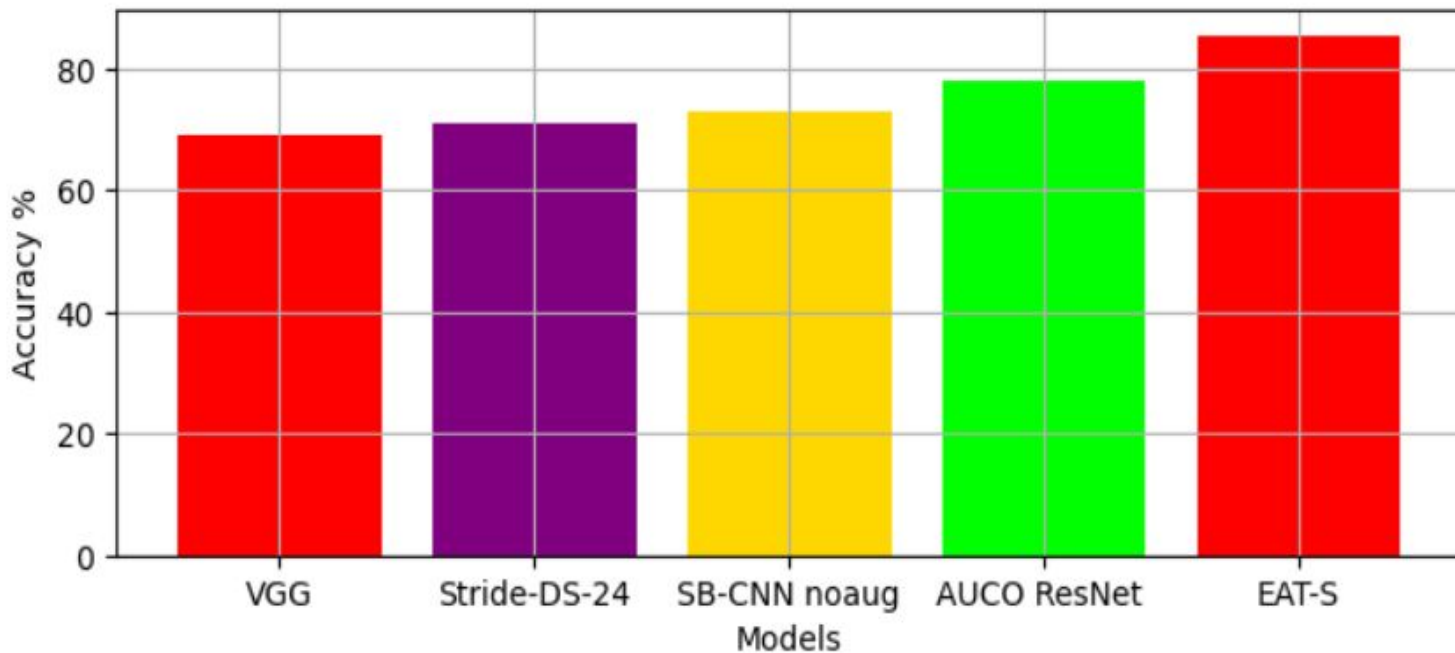




Zaključak

- Pre-trenirani VGG je ubedljivo najbolji model
- Na VGG nismo utrošili vreme na implementaciju
- Vreme potrebno za obučavanje VGG nije značajno veće od vremena potrebnog za obučavanje CNN
- Jednostavan za korišćenje

Poređenje sa dostupnim modelima, bez korišćenja dodatnih podataka



- EAT - S: <https://paperswithcode.com/paper/end-to-end-audio-strikes-back-boosting>
- AUCC ResNet:
<https://paperswithcode.com/paper/aucc-resnet-an-end-to-end-network-for-covid>
- SB - CNN noaug:
<https://paperswithcode.com/paper/deep-convolutional-neural-networks-and-data-1>
- Stride - DS - 24:
<https://paperswithcode.com/paper/environmental-sound-classification-on>