

Prepoznavanje muzičkog instrumenta i tona

Autori : Jovana Kostreš i Marija Ćurčić
Fakultet tehničkih nauka, Univerzitet u Novom Sadu

Uvod

Cilj ovog projekta je rešavanje problema prepoznavanja tona i muzičkog instrumenta kojim je ton odsviran. Vršiti se prepoznavanje 12 vrsta tonova i 7 vrsta instrumenata na osnovu audio zapisa koji sadrže samo jedan ton. Tonovi zapisa se određuju nezavisno od oktave kojoj pripadaju. Prepoznaju se tonovi u muzičkoj lestvici: C, C#, D, D#, E, F, F#, G, G#, A, A#, B. Prepoznaju se sledeći muzički instrumenti: harmonika, violina, violončelo, viola, kontrabas, rog, B klarinet.

Motivacija

Rešenje problema prepoznavanja muzičkih instrumenata i tonova na osnovu audio zapisa može se upotrebiti u muzičkoj transkripciji. Ovo rešenje se može proširiti tako da se vrši i prepoznavanje trajanja nekog tona, oktave kojoj pripada i slično.

Metodologija

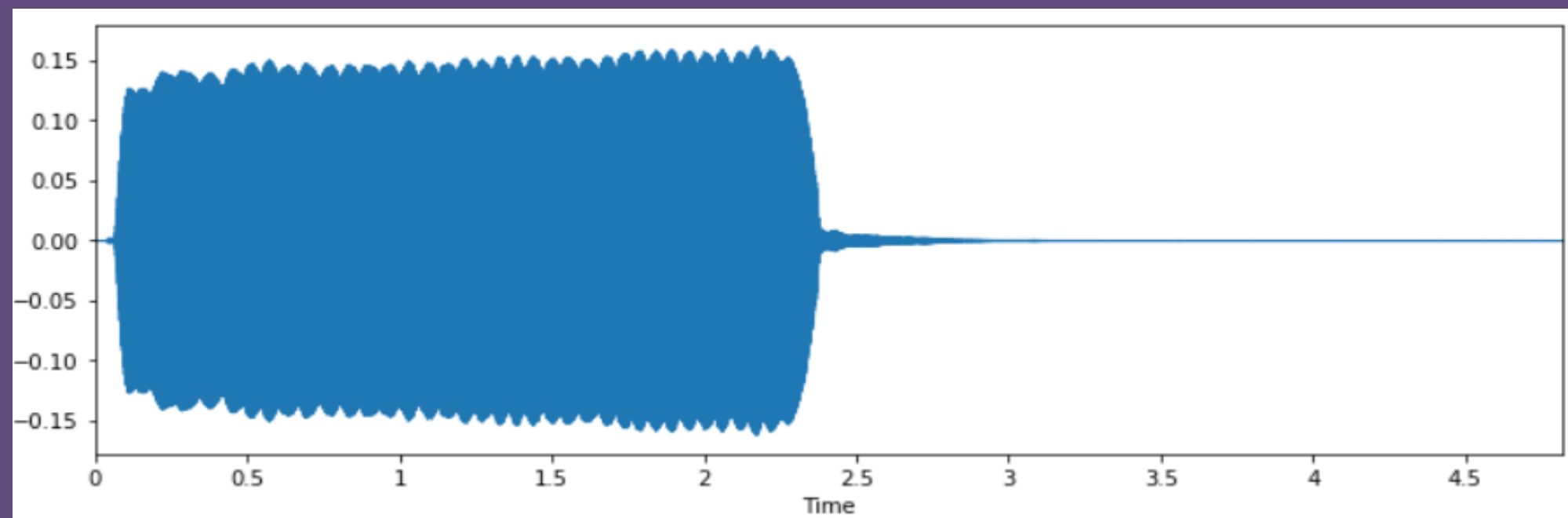
Za obradu podataka korišćena je biblioteka LibROSA i njenje metode za učitavanje, isecanje zvuka i kreiranje spektrograma. Rešavanje problema se vrši kroz dva dela:

- Prepoznavanje muzičkog instrumenta pomoću izmenjene AlexNet mreže. Skup podataka je podeljen na skupove za trening, validaciju i test u odnosu 80:10:10. Efikasnost se određuje korišćenjem metrika accuracy i loss.
- Prepoznavanje tona se vrši na identičan način kao i za muzički instrument sa razlikom broja krajnjih klasa.

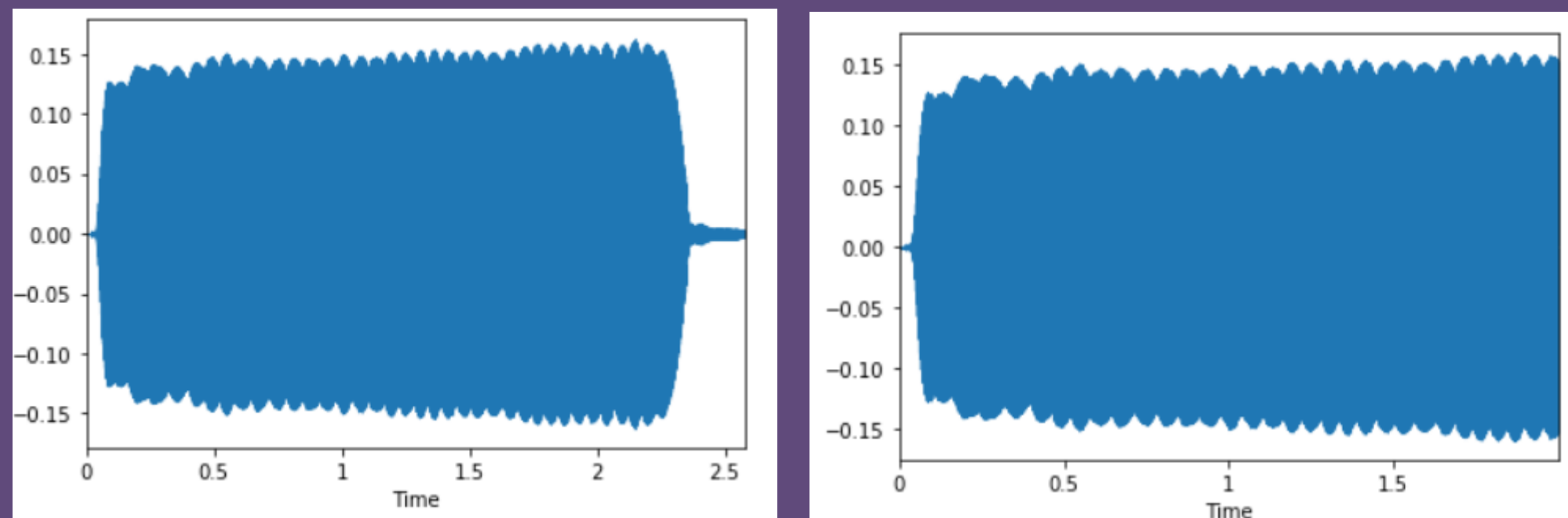
Skup podataka i obrada podataka

Podaci su preuzeti sa sajtova TinySOL i UIOWA:MIS. Ukupan broj podataka koji su korišćeni za izradu podataka jeste 1937.

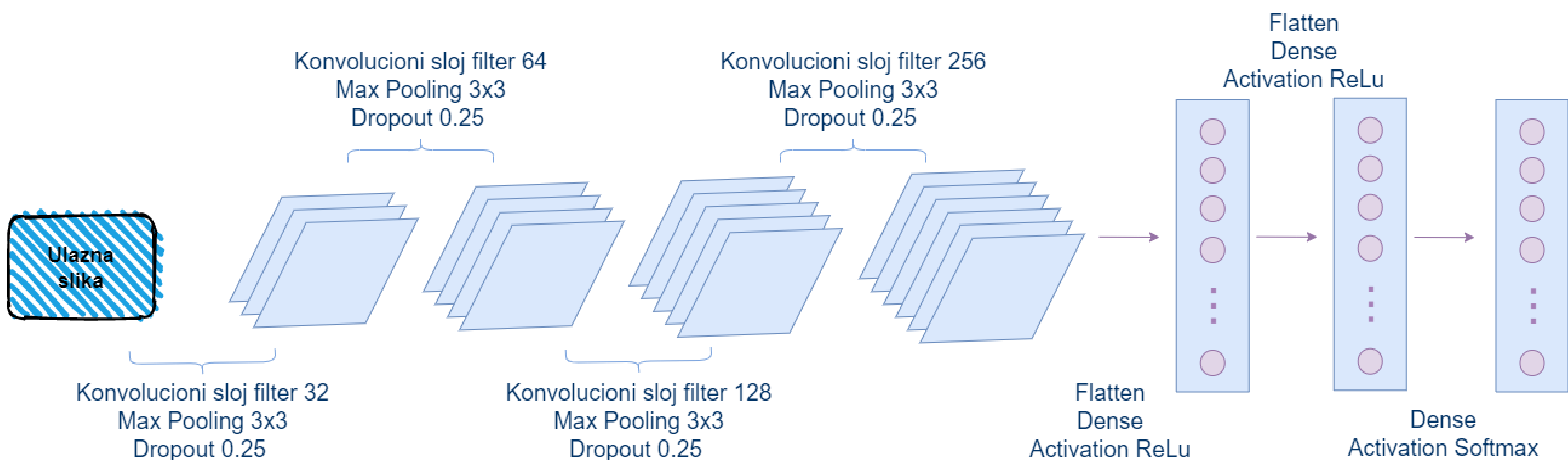
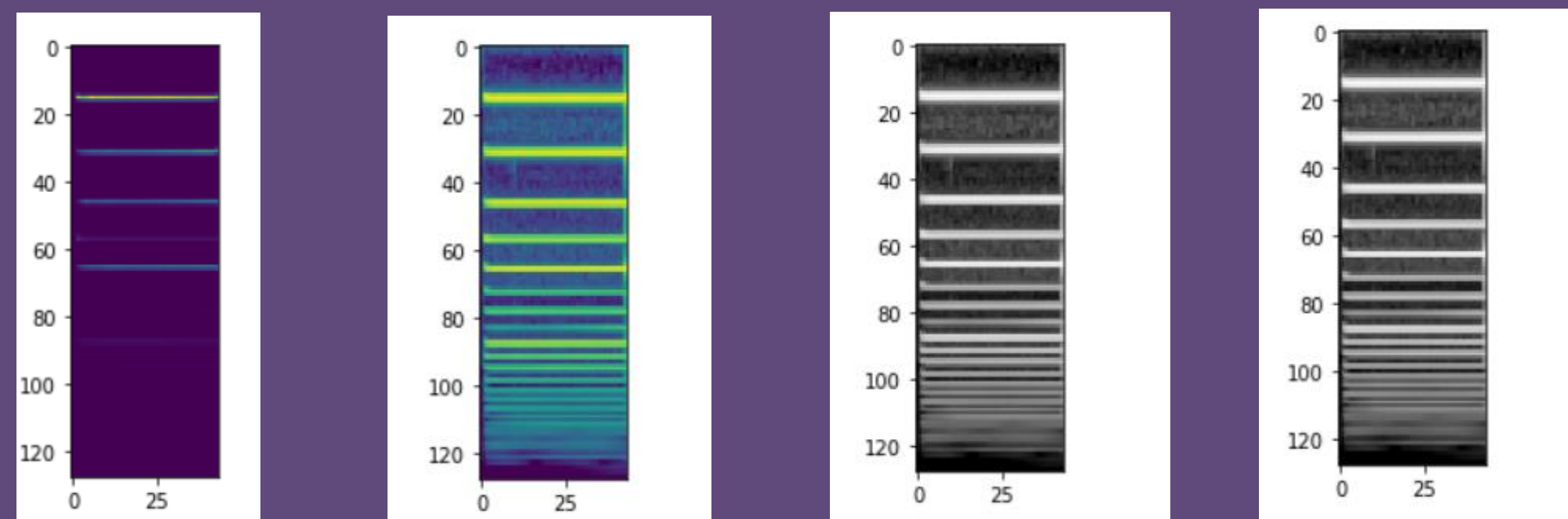
Podaci su prvo učitani u njihovom originalnom obliku sa stopom uzorkovanja 22050Hz.



Zatim se kreira iseca tišina iz učitanoz zvuka uklanjanjem zvuka ispod 30dB. Zatim se zvuk seče na dužinu od dve sekunde kako bi svi podaci bili iste veličine.



Od obrađenog zvuka se dalje kreira Melov spectrogram i radi se dalja obrada spektrograma konvertovanjem amplituda u decibele. Zatim se slika konvertuje u nijanse sivih i na kraju se radi njena normalizacija.



Arhitektura rešenja

Arhitekturu rešenja predstavljaju dve CNN mreže koje su nastale kao izmena AlexNet mreže. Pre puštanja slika u mrežu one su skalirane na 128 x 44 px prethodnim sečenjem zvuka.

Konvoluciona mreža za prepoznavanje tona se sastoji od 4x ponovljenog konvolucionog bloka, Flatten sloja, 2 x Fully connected sloja i Softmax aktivacije.

Konvolucioni blok se sastoji od:

- Konvolucionog sloja
- Batch normalizacije
- 3 x 3 Max pooling-a
- 0.25 Dropout-a

Rezultati

Mreže su trenirane u 60 epoha.

Neural network	Training accuracy	Test accuracy
CNN for instruments	0.97	0.94
CNN for pitch	0.94	0.93

Zaključak

Korišćenjem napomenute mreže postignuti su zadovoljavajući rezultati, međutim uvek ima mesta za poboljšanje. Proširenje skupa podataka je jedan od načina za moguće poboljšanje rada mreže. Takođe, postoji mogućnost da bi promena arhitekture same mreže donela bolje rezultate, međutim, to bi verovatno iziskivalo i povećanje kapaciteta hardvera.