



MATEMATIČKI FAKULTET

SEMINARSKI RAD IZ RAČUNARSKE INTELIGENCIJE

Analiza forhenda

Autor:
Mihailo Dedić

Profesor:
Prof. Dr. Vladimir Filipović

Asistent:
Stefan Kapunac

25. jun 2024.

Sadržaj

1 Uvod i opis problema	2
2 Pretprocesiranje	3
2.1 Detekcija zglobova	3
2.2 Normalizacija pozicija zglobova	4
2.3 Aproksimacija forhenda skupom krivih	5
2.4 Formiranje normalizovanih krivih	6
2.5 Formiranje skupa podataka	7
3 Modeli i rezultati	9
3.1 Potpuno povezana neuronska mreža	9
3.2 Konvolutivna neuronska mreža sa 1 kanalom boje	10
3.3 Konvolutivna neuronska mreža sa 3 kanala boje	13
4 Zaključak	17

1 Uvod i opis problema

Popularizacija tenisa sa pocetka 21. veka doprinela je do tada neviđenom nivou profesionalizma. Mnogi od tenisera počeli su da se obraćaju za pomoć data analiticarima radi analize i unapređivanja njihove igre. Sa druge strane, intenzivnim napretkom veštačke inteligencije i računarskog vida došli smo u situaciju da je moguće vršiti analize raznih aspekata igre. Pokrenuti su mnogi *startup-i* koje se bave analizom kvaliteta udaraca (npr. *Tennis Insights*), zatim uvidom u tehniku i strategiju (npr. *Swing Vision*)...

Takođe su napisani brojni naučni radovi na ovu temu u zadnjih pet godina. Možda i najkompletniji rad na ovu temu je *Learning Physically Simulated Tennis Skills from Broadcast Videos* [1] od grupe autora predvođene Haotianom Zhangom. U ovom radu naučnici su se bavili kreiraranjem avatara koji je u mogućnosti da nauči udarce samo na osnovu videa teniskih mečeva. (Ovaj rad je i nagradjen...)

Ovaj rad je jednim delom inspirisan sledecim radom: *Stroke Comparison between Professional Tennis Players and Amateur Players using Advanced Computer Vision*[2] od naučnika iz Tokija. Glavna tema ovog rada bila je analiza oscilacija servisa kod amatera i profesionalnih igrača. Zaključeno je da profesionalci imaju mnogo manje oscilacije kod između dva svoja servisa nego amateri. Neke od tehnika normalizacije, iz prethodno navedenog rada, upotrebljene su delom i u ovom projektu.

U ovom radu će glavna tema biti klasifikacija forhenda igrača na osnovu datih video snimaka treninga. Želja je da se otkriju razlike u forhendima samo na osnovu pokreta igrača, nezavisno od video snimka, ugla kamere i pozicije igrača. Ljubiteljima tenisa lako je da prepoznaju tehniku svakog od vrhunskih igrača, pa ćemo stoga ovde probati da to naučimo model da to uradi.



(a) Prikaz Alkaraza pri udaranju forhenda



(b) Prikaz Sinera pri udaranju forhenda

Slika 1: Prikaz Alkaraza i Sinera pri udaranju forhenda

2 Preprocesiranje

Glavnu prepreku u izradi je predstavljalo preprocesiranje, to jest kako doći od običnih video snimaka treninga igrača do skupa podataka koji se može proslediti modelu. U nastavku će detaljnije biti opisan svaki od segmenata preprocesiranja.

2.1 Detekcija zglobova

Video snimci predstavljaju smenu određenog broja slika nekog kvaliteta u sekundi - takozvani frame-rate(fps). Za analizu ćemo koristiti samo one video snimke koji su kvaliteta 4k sa 60fps. Izabrano je 3 videa - po jedan za svakog od sledećih igrača - Grigora Dimitrova, Janika Sinera i Karlosa Alkaraza. Svaki od videa pomoću alata ffmpeg je podeljen na frejmove. Zatim smo rucno tražili trenutke udaraca pojedinačnih forhenda - ovaj frejm smo nazvali *hitting frame*. Uzeto je i 15 frejmova koji su mu prethodili, i 10 frejmova posle njega. Dakle, od jednog udaraca dobili smo 26 frejmova. Ponavljujući ovaj postupak, uspeli smo da izdvojimo 43 forhenda od Dimitrova i po 48 forhenda od Sinera i Alkaraza. Treba napomenuti da smo se trudili da forhendi budu relativno oko sredine terena, pri normalnom nameštanju igrača (nismo obraćali pažnju na nezgodne lopte kada je igrač morao da pravi brze promene kretanje kako bi došao do njih, jer oni izazivaju problem inercije kojim se nećemo baviti u ovom radu). Svaki od ovih forhenda je zatim provučen kroz *Alpha Pose* sistem (uz korišćenje Pose Flow-a)[3] pomoću koga smo dobili pozicije 17 ključnih ljudskih zglobova u svakom od 26 frejmova:



Slika 2: Prikaz zglobova jednog igrača u jednom frejmu

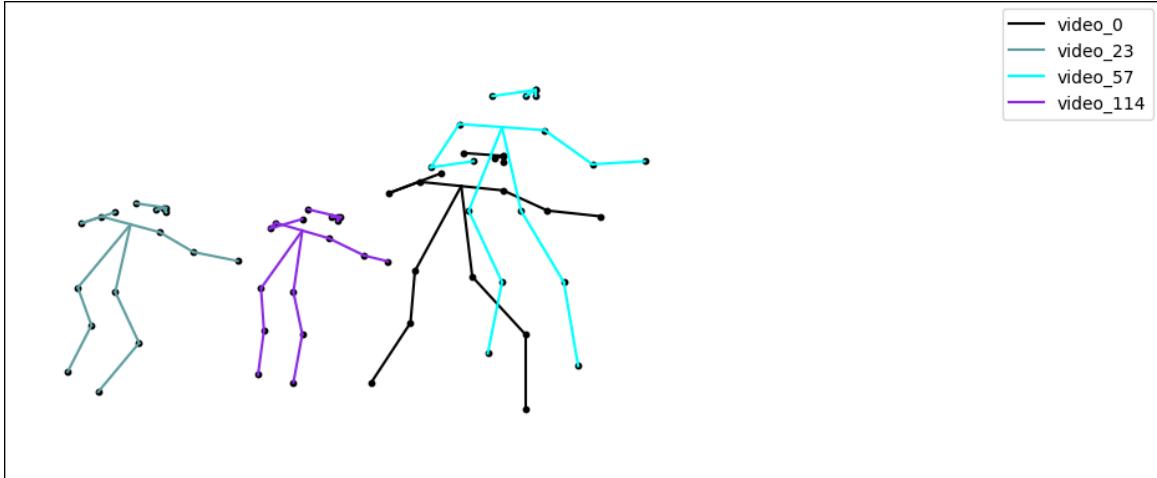
Kao što se vidi na slici, sistem je izdvojio zglove za većinu aktera snimka. Koriste se samo zglove igrača bližeg kamери, koji je detektovan tako što se uzme minimalna vrednost na y-osi desnog zgloba noge. Napominje se još da je sistem veoma dobro radio u slučaju Sinera i Alkaraza, ali je pratio dosta grešaka u slučaju Dimitrovog videa.



Slika 3: Prikaz lošeg rada sistema u slučaju Dimitrova

2.2 Normalizacija pozicija zglobova

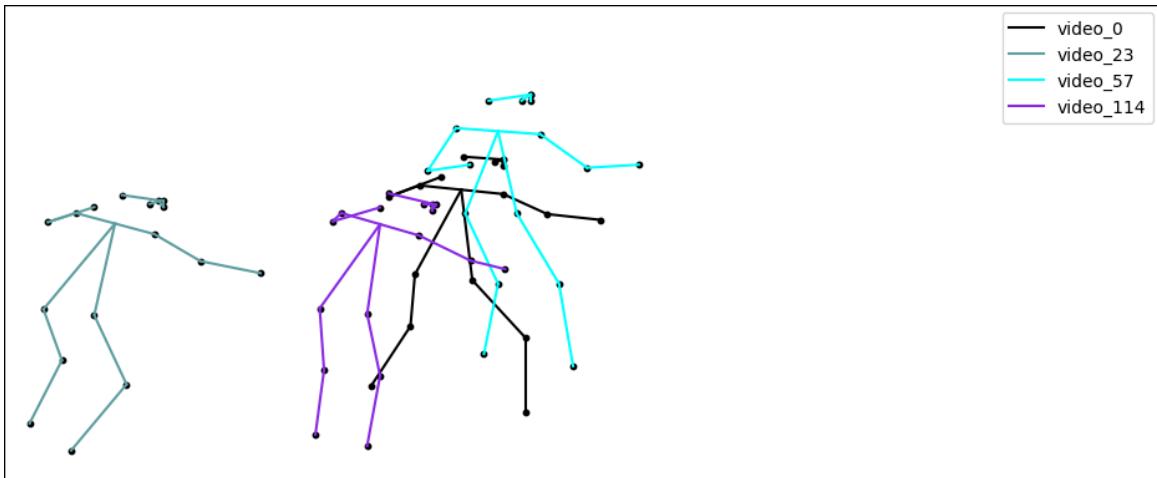
Svaki od korišćenih video snimaka sa razlikuje jedan od drugog prema ugлу snimanja. Štaviše, svaki udarac se dešava na različitoj poziciji terena, a takođe postoji i razlika u građi izmedju igrača. Jasno je da će biti potrebno predstaviti ih u nekom sličnom formatu kako bi mogla da se vrše poređenja.



Slika 4: Prikaz bez normalizacije

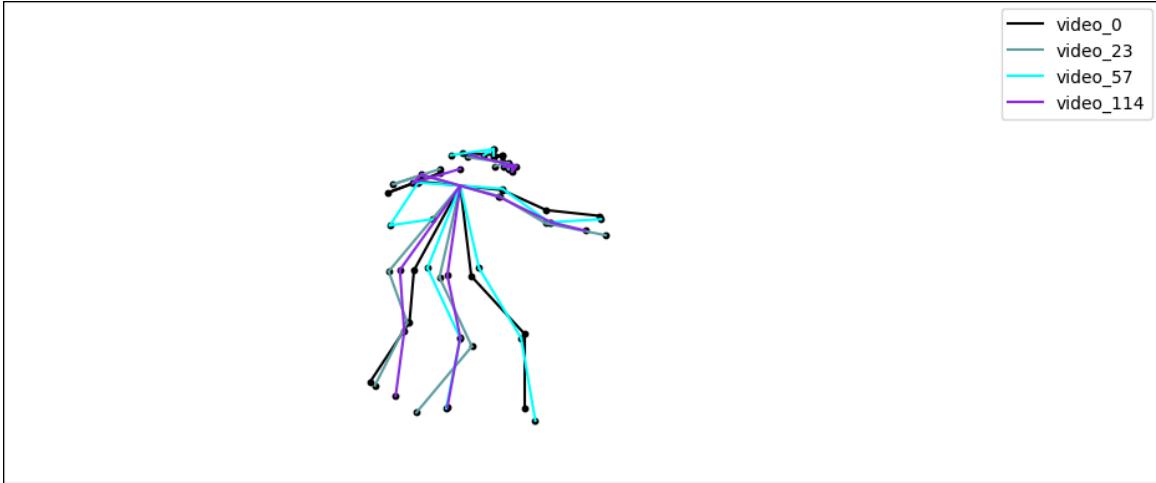
Normalizacija mora biti vršena u odnosu na neki referentni skup zglobova (skeleton igrača). Umesto da se računa srednja vrednost pozicija zglobova za sve forhende, što bi bilo računski zahtevno, odlučeno je da se koristi jedan referentni forhend koji je subjektivno procenjen kao dobro postavljen na terenu i u odnosu na kameru. Ovaj forhend će se u daljem tekstu označavati kao inicijalni forhend.

Prvi korak normalizacije predstavlja skaliranje. Za svaki forhend, faktor skaliranja ćemo odrediti tako što ćemo podeliti dužinu ramenog pojasa naseg forhenda u *hitting* frejmu, sa dužinom ramenog pojasa inicijalnog forhenda u *hitting* frejmu. Centar skaliranja ćemo dobiti tako što izdvojimo po dve najekstremije pozicije skeleta na x i y osi u *hitting* frejmu, i nađemo njihov prosek za svaku od osa. Skaliranje sa ovim faktorom i centrom ćemo zatim primeniti na svaki od zglobova u svakom od frejmova naseg forhenda.



Slika 5: Prikaz sa skaliranjem

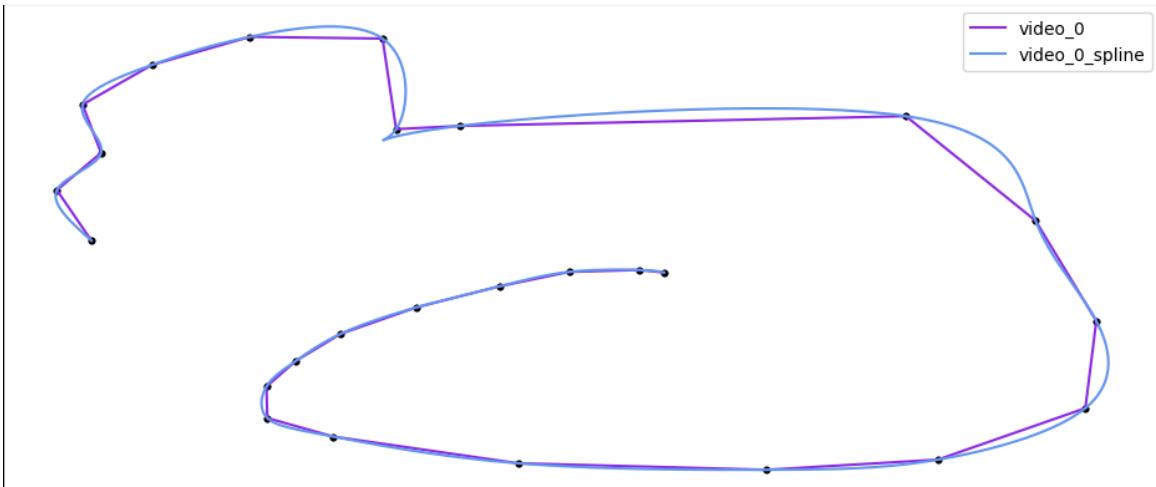
Drugi korak normalizacije predstavlja translaciju. Svaki forhend ćemo pomeriti tako da poziciju centra ramenog pojasa u *hitting* frejmu poklopimo sa pozicijom centra ramenog pojasa inicijalnog forhenda.



Slika 6: Prikaz sa skaliranjem i translacijom

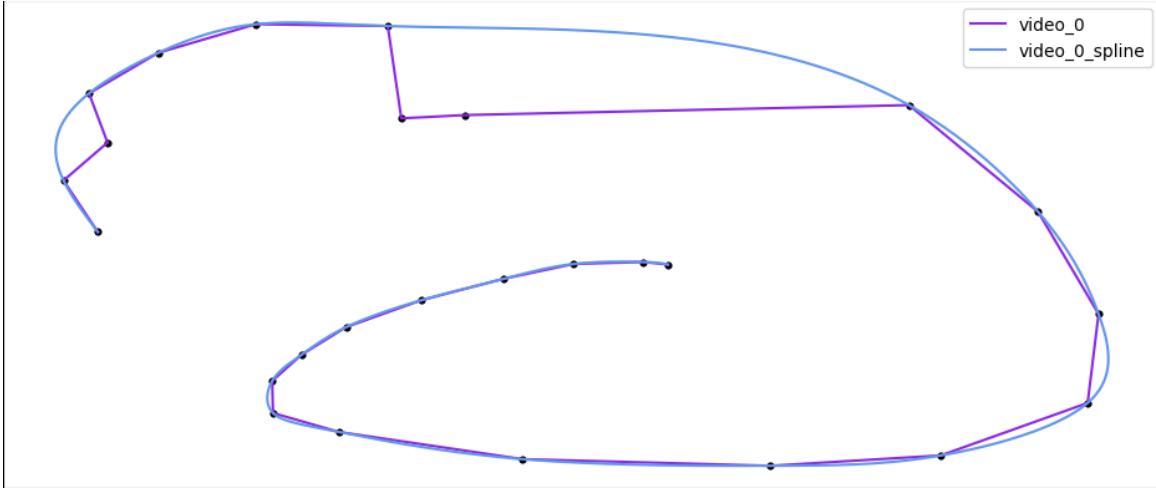
2.3 Aproksimacija forhenda skupom krivih

Želja je da se nekako aproksimira konkretni forhend. Dakle, pitanje je kako predstaviti video, tj. 26 frejmova, kao jedan podatak. Posmatrajmo pozicije zglobova kroz frejmove posebno za svaki zglob. Plan je da se provuče kriva koja će aproksimirati putanje tog zgloba kroz ceo video. Koristi se CubicSpline iz Scipy biblioteke sa parametrom `bc_type=natural...`



Slika 7: Osnovna aproksimacija

Putanje svakog od zglobova, ili bar onih koji prelaze najveće rastojanje kao što su desni zglob i desni lakat očekivaćemo da budu kao pravilne krive. Kao što vidimo na slici iznad kod tačke u 19. frejmu imamo naglu promenu pravca kretanja krive što predstavlja neku grešku. Utvrđeno je da su te greške ustvari pogrešne detekcije pozicija zglobova od strane AlphaPose sistema. Sa ovim greskama ćemo se boriti na sledeći način:
Kod forhenda se vrši rotacija tela uлево (odnosno rotacija tela pozitivnim matematičkim smerom). Posmatraju se svake 3 vezane tačke počevši od prvog frejma, tj. pozicije zgloba u 3 vezana frejma. Od njih se formira trougao i računa njegovua orijentacija. Ako je njegova orijentacija negativna, srednja od tih tacaka se označava kao *outlier* i izbacuje se iz daljeg razmatranja. Analogni princip se radi počevši od poslednjeg frejma samo sto će *outlier* sada biti srednja tačka trougla koji je pozitivno orijentisan. Na ovaj način se dobija sledeća aproksimacija putanje zgloba (Kriva se formira samo od tačaka koje nisu izbačene i njihovih odgovarajućih frejmova):



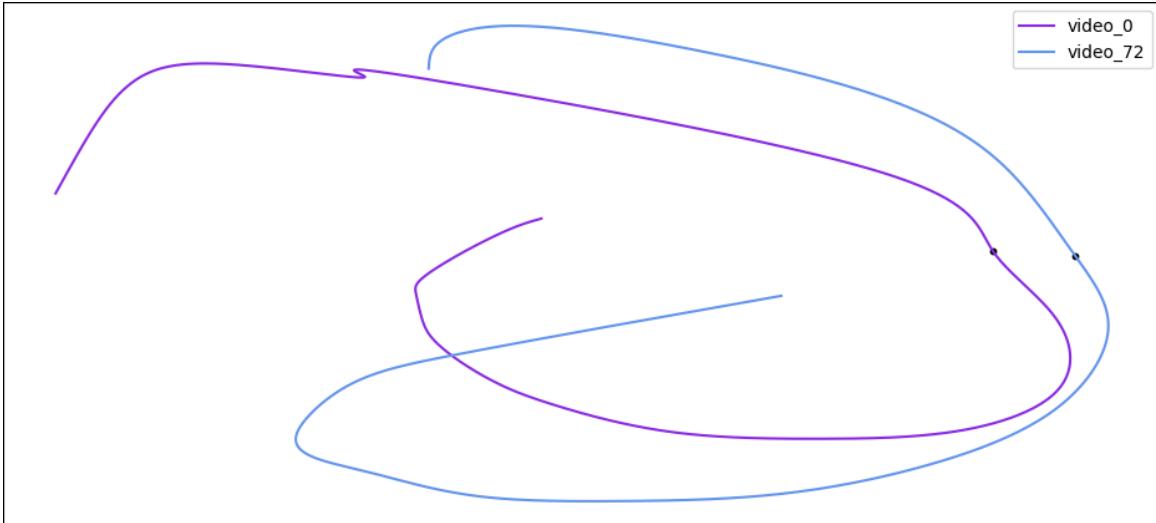
Slika 8: Aproksimacija sa kompletnim uklanjanjem outlier-a

2.4 Formiranje normalizovanih krivih

Svaka od krivih se translira za rastojanje njene tačke u *hitting* frejmu od odgovarajuće tačke inicijalne krive (krive za inicijalni forhend) u *hitting* frejmu.

Početni i završni frejmovi mogu ometati analizu, jer se dosta razlikuju kod igrača, a nemaju preterani uticaj na udarac, tako da će se posmatrati kriva od 6. zaključno sa 21. frejmom gde je 16. frejm - *hitting* frejm. Inicijalna kriva će se zatim parametrizovati za te frejmove i odgovarajuće tačke u njima, te će biti nazvana normalizovanom inicijalnom krivom.

Krive se mogu dosta razlikovati po dužini, npr. zbog brzine udarca - ako je udarac brži biće uhvaćeno dosta više pokreta za fiksiran broj frejmova od tehnički istog udarca, koji je samo sporije udaren.

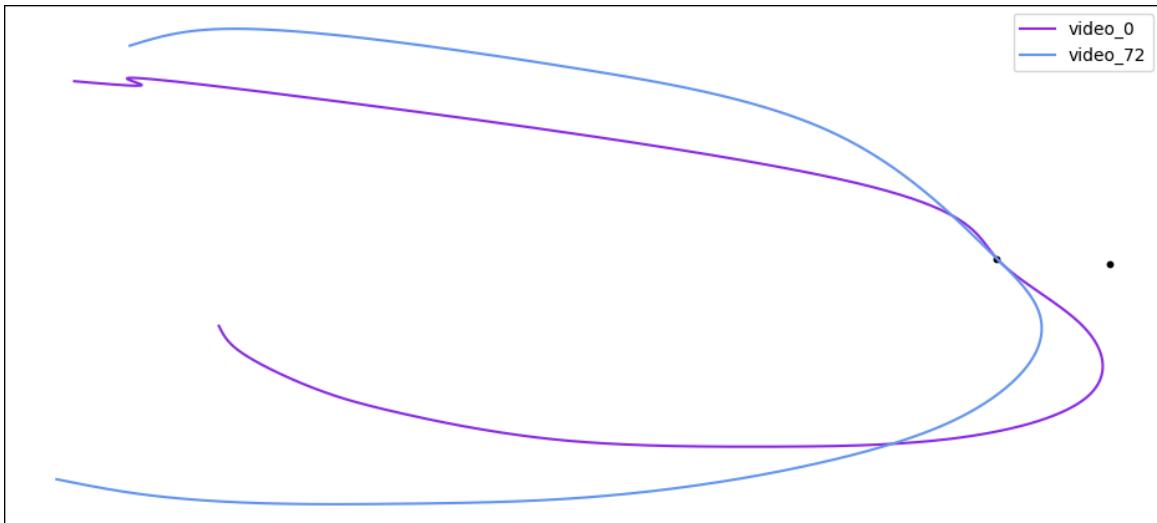


Slika 9: Poređenje 2 krive

Cilj je da sve odgovarajuće krive budu slične dužine. Zbog toga se za ostale krive vrši sledeći postupak:

1. Određuje se t_{\max} . Kreće se od *hitting* frejma unapred i računa se na nenormalizovanoj krivi tačka u kojoj je rastojanje od *hitting* frejma isto kao i odgovarajuće rastojanje kod normalizovane inicijalne krive i ona se obeležava kao t_{\max} (ako je taj udarac brži nego inicijalni $t_{\max} < 21$, inače je $t_{\max} > 21$ što je i jedini razlog sto zadržavamo zadnjih 5 frejmova (isto tako i za prvih 5))
2. Određuje se t_{\min} analogno. Kreće se od *hitting* frejma unazad i računa na nenormalizovanoj krivi tačka u kojoj je rastojanje od *hitting* frejma isto kao i odgovarajuće rastojanje kod normalizovane inicijalne krive i ona se obeležava kao t_{\min} (ako je taj udarac brži nego inicijalni $t_{\min} > 6$, inace je $t_{\min} < 6$)
3. Određuje se 16 tačaka na istom rastojanju između t_{\min} i t_{\max} koje će predstavljati isti vremenski trenutak kao i 16 frejmova u normalizovanoj inicijalnoj krivi i racunamo vrednost krive u svakoj od njih

4. Formira se normalizovana kriva na osnovu ovih 16 tačaka i vrednosti u njima



Slika 10: Poredjenje 2 krive posle normalizacije

2.5 Formiranje skupa podataka

Trenutno je napravljeno 17 normalizovanih kriva koje zajedno predstavljaju forhend igrača. Za pocetak izbacujemo 5 zglobova glave, jer je pozicija glave tokom udarca nepromenljiva i ne određuje udarac. Posmatrajmo putanje tih zglobova:



Slika 11: Komplet igrač

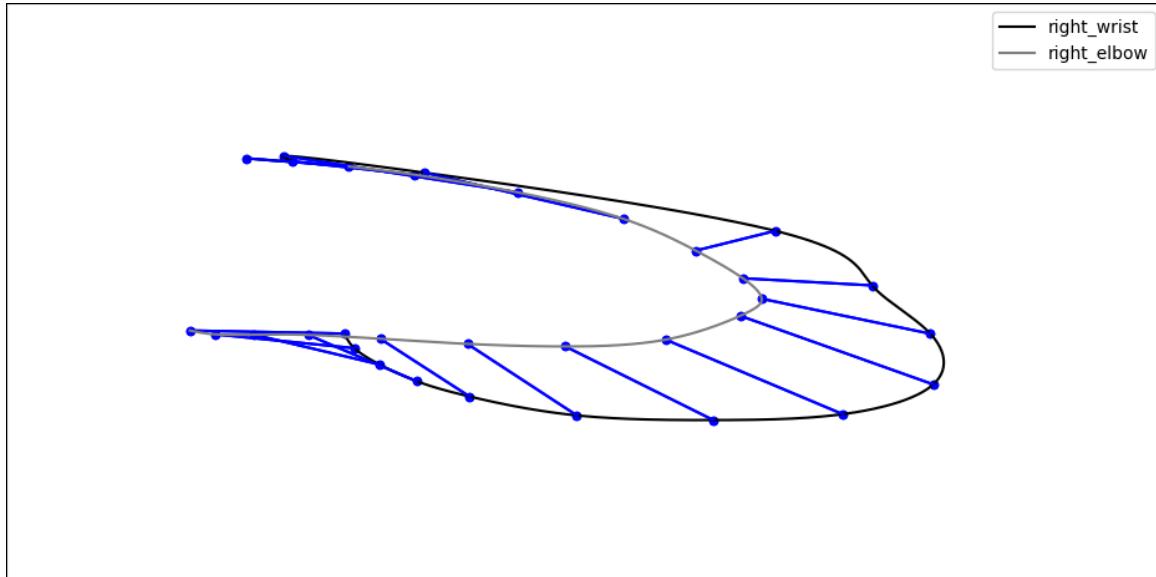
Ono što se može videti za ostale zglove je da dosta zavise od tipa dolazne lopte, npr. za kraću loptu igrač će ispružiti levu nogu napred i doći do određene pozicije, a za dužu će doći u skoro paralelan stav, što će rezultovati skroz drugom pozicijom zglobova. Ono što ne bi trebalo da zavisi od pozicije igrača (ili bar da ima dovoljno mali uticaj) jeste desni zglob i desni lakat. Oni samo služe da odrade putanju reketa. Zato će oni biti u fokusu naše analize.

Dakle cilj je da se odredi forhend igrača preko veze između krive desnog zglobova i desnog lakteta. Ovo je urađeno na sledeći nacin:

1. Prebacuju se kontrolne tacke ovih normalizovanih krivih u prostor $[0, 1]$ kako ne bi bilo potrebe da se vršimo posle standardizacija
2. Računa se rastojanje između desnog lakteta i desnog zglobova u svakoj kontrolnoj tački

3. Rastojanju se daje predznak minus akko desni lakat seče desni zglob posle 8. frejma, s tim da ako ga ponovo presece predznak se briše

Na ovaj način se dobije 16 rastojanja koji će u potpunosti određivati jedan forhend u skupu podataka.



Slika 12: Zglob - Lakat

3 Modeli i rezultati

Problem se rešava pomoću tri modela. Koristi se jedan model potpuno povezanih neuronskih mreža, jedan model konvolutivnih neuronskih mreža sa jednim kanalom boje i jedan model konvolutivnih neuronskih mreža sa tri kanala boje.

3.1 Potpuno povezana neuronska mreža

Posmatrati prvo slučaj kada se porede sva 3 igrača: Skup podataka će se sastojati od 139 instanci, od kojih će svaka imati 16 atributa rastojanja izračunatih u prethodnom poglavlju. Svaka instanca će imati i svoju klasu:

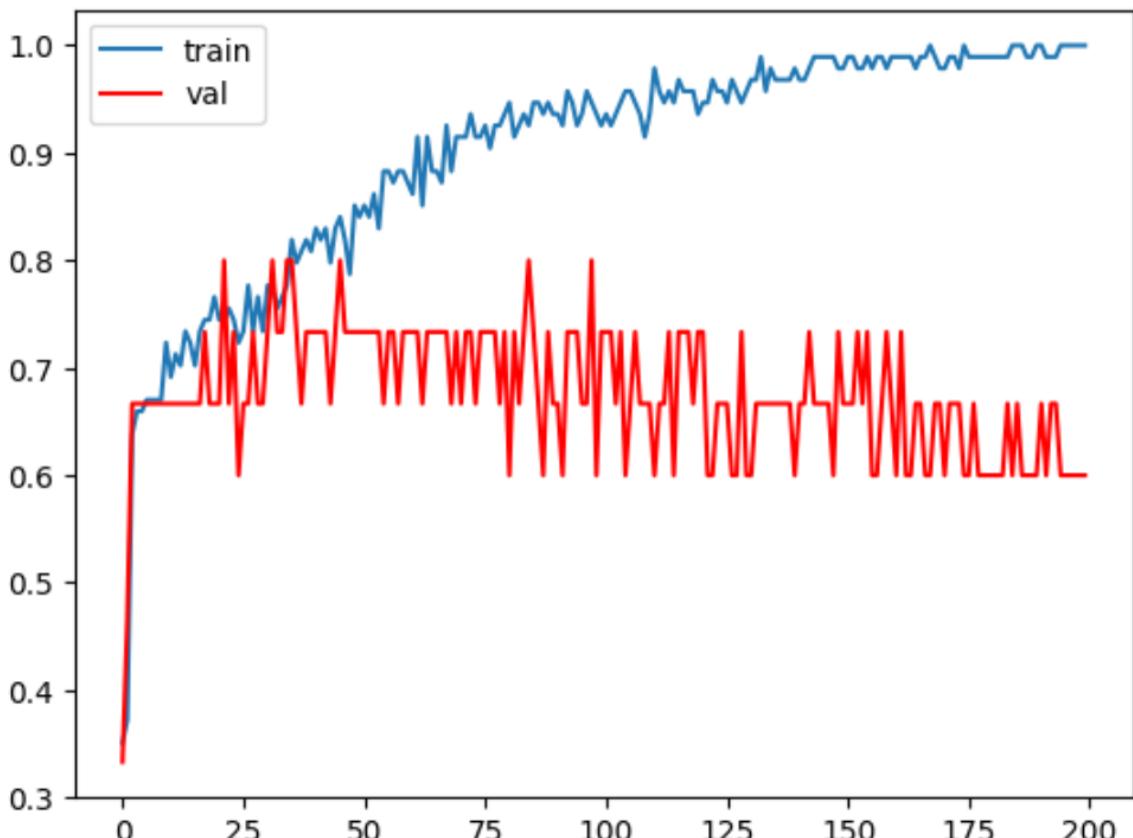
- 0 - Dimitrov,
- 1 - Siner,
- 2 - Alkaraz.

Konstruiše se model sa sledećom arhitekturom:

```
Net(  
    (layer1): Linear(in_features=16, out_features=64, bias=True)  
    (layer2): Linear(in_features=64, out_features=32, bias=True)  
    (layer3): Linear(in_features=32, out_features=16, bias=True)  
    (layer4): Linear(in_features=16, out_features=3, bias=True)  
    (activation): ReLU()  
)
```

Slika 13: Arhitektura potpuno povezanog modela

Model se trenira 200 epoha sa tim da će se za krajnji izabratiti onaj model koji je u 2 vezane epohe prvi ostvario najbolji rezultat tačnosti na validacionom skupu. Kao što se vidi sa slike, to se ovde postiže u epohi 35:

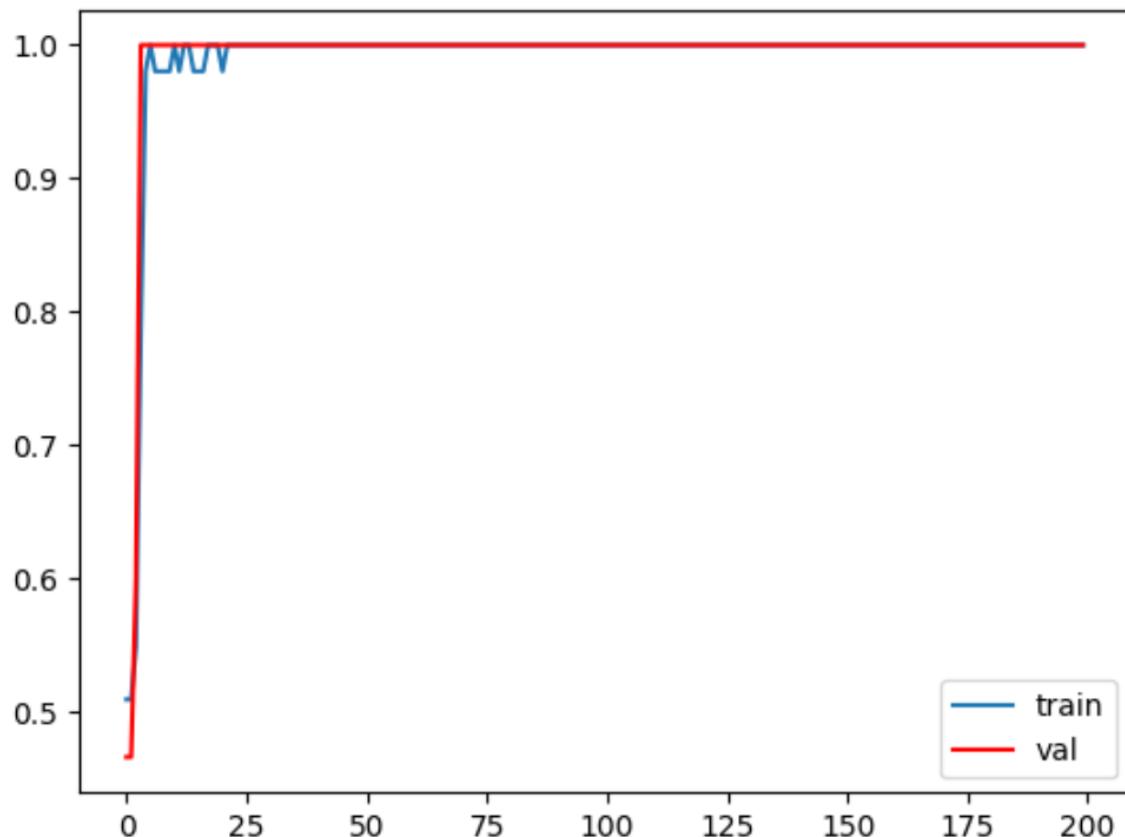


Slika 14: Tačnost na trening i validacionom skupu kroz epohe

Tačnost na test skupu iznosi 73.33%.

Takođe, konstruiše se jedan test skup koji se sastoji od 10 forhenda Dimitrova i po 7 forhenda Sinera i Alkaraza, a koji svi pripadaju drugaćijim videima od onih na kojima su trenirani. Vidi se da je tačnost ostala slična - 75%, što znači da model dobro uopštava i na forhende sa drugih videa, što je veoma značajno za neku praktičnu primenu ovog projekta.

Ono što je takođe zanimljivo je da ima dosta problema sa videom Dimitrova na kome je izvršen trening, tako da se može posmatrati i samo skup Sinera i Alkaraza (uz samo malu promenu modela). Vidi se sa sledeće slike da se bira već epoha 4:



Slika 15: Tačnost na trening i validacionom skupu kroz epohe

Dobijaju se sledeći rezultati na glavnom test skupu i dodatnom test skupu:

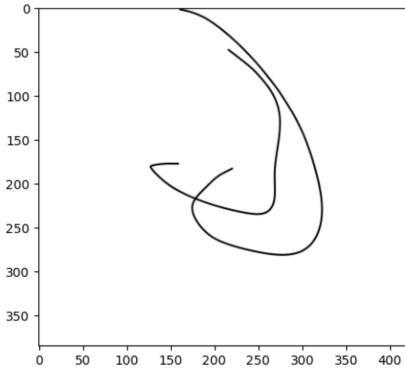
glavni test skup - 96.67%

dodatni test skup - 100%

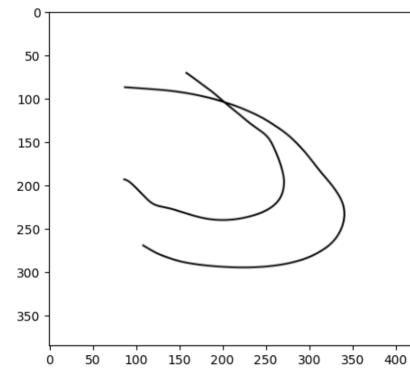
Zaključak je da je model, skoro u potpunosti, u stanju da razlikuje forhende Sinera i Alkaraza.

3.2 Konvolutivna neuronska mreža sa 1 kanalom boje

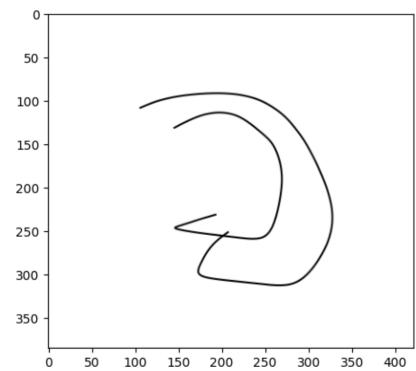
Ponovo će se posmatrati prvo slučaj kada se porede sva 3 igrača: Skup podataka će se sastojati od 139 instanci, koje će predstavljati crno-belu sliku putanje desnog zgloba i desnog lakta. Na slici dole se vidi po jedna instance za svaku od 3 klase:



(a) Primer forhenda Dimitrova



(b) Primer forhenda Sinera



(c) Primer forhenda Alkaraza

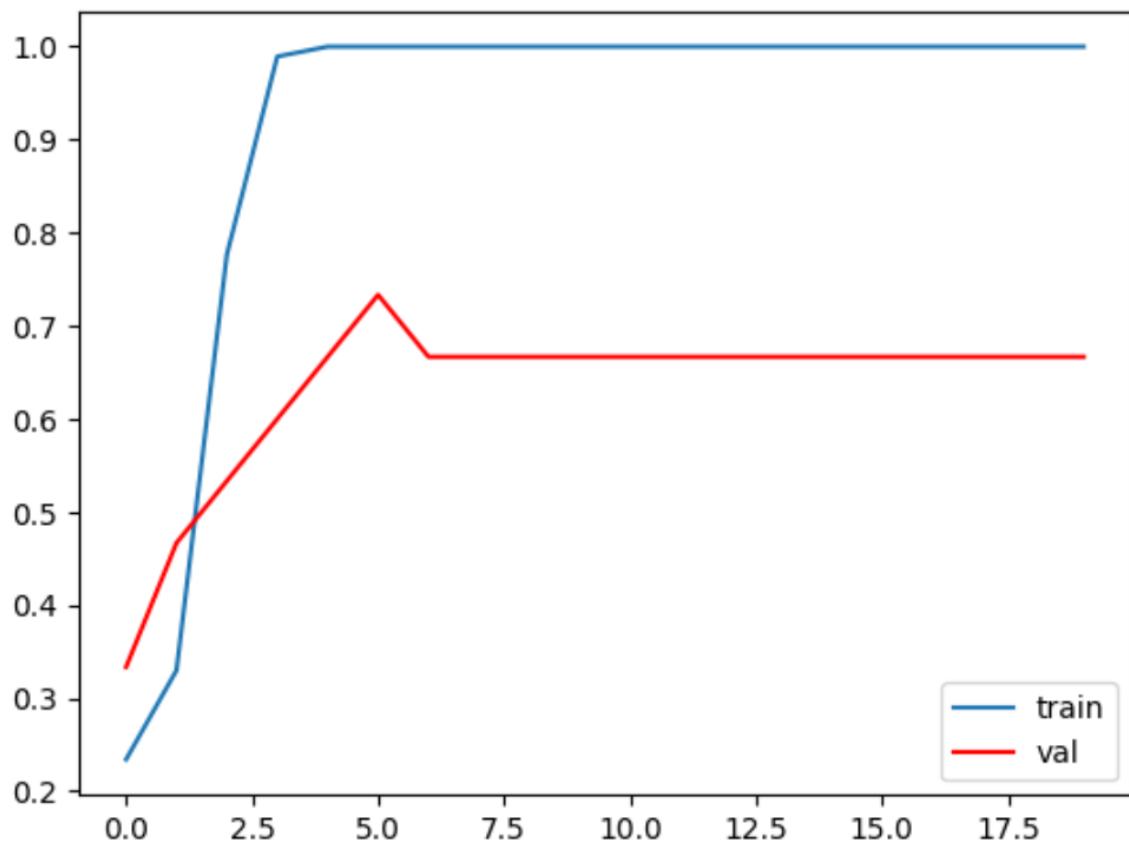
Slika 16: Prikaz instanci skupa

Arhitektura modela:

```
ConvNet(
    (layer1): Conv2d(1, 16, kernel_size=(3, 3), stride=(1, 1))
    (layer2): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1))
    (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (flatten): Flatten(start_dim=1, end_dim=-1)
    (activation): ReLU()
    (layer3): Linear(in_features=312832, out_features=128, bias=True)
    (fc): Linear(in_features=128, out_features=3, bias=True)
)
```

Slika 17: Arhitektura konvolutivnog modela sa 1 kanalom

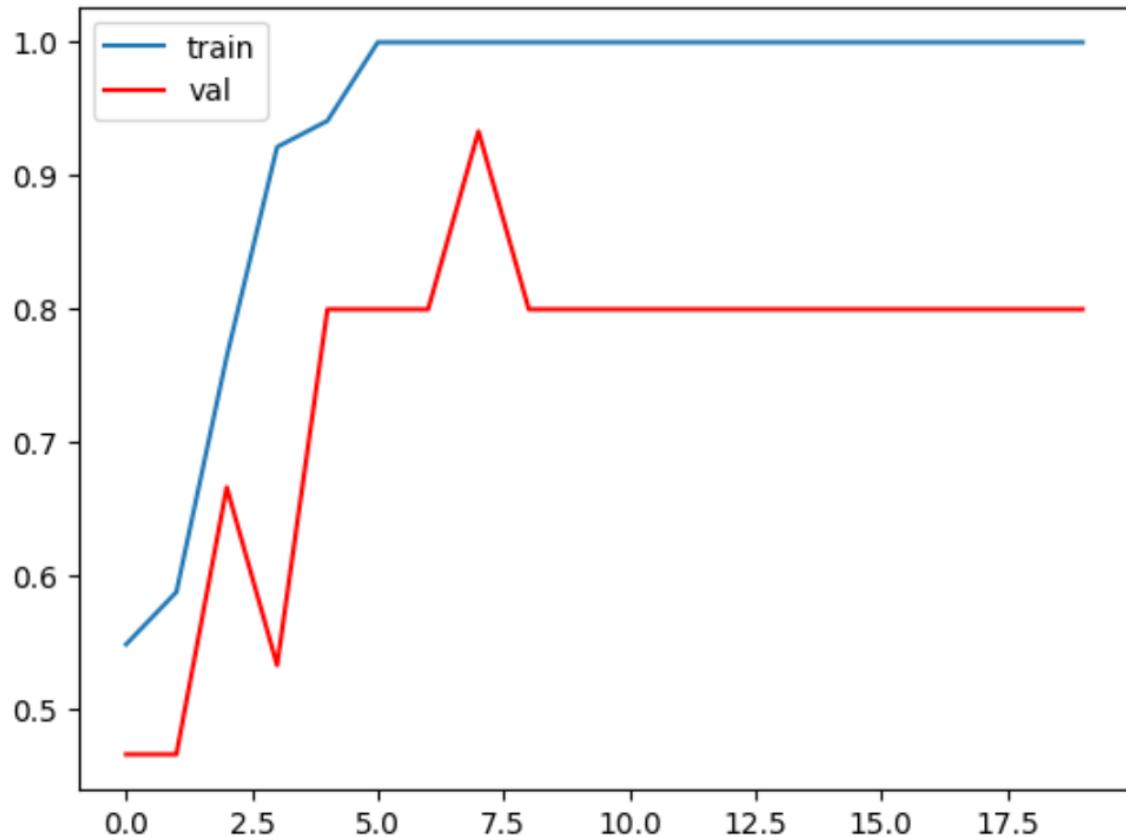
Ovaj model će se trenirati 20 epoha, a kao što se vidi sa slike, najbolje karakteristike se ostvaruju u 7. epohi:



Slika 18: Tačnost na trening i validacionom skupu kroz epohe

Rezultati dobijeni na glavnom test skupu i dodatnom test skupu su:
glavni test skup - 60%
dodatni test skup - 62.5%

Što se tiče skupa Alkaraza i Sinera, najbolji rezultat se postiže u 9. epohi :



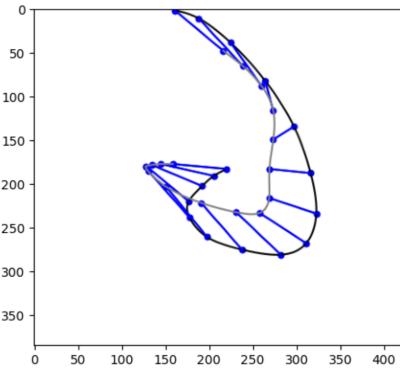
Slika 19: Tačnost na trening i validacionom skupu kroz epohe

Ovde se dobijaju sledeći rezultati tačnosti:

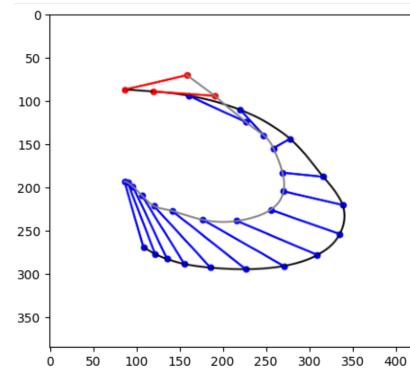
glavni test skup - 90%
dodatni test skup - 85.71%

3.3 Konvolutivna neuronska mreža sa 3 kanala boje

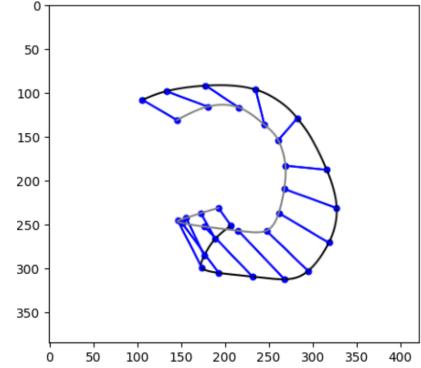
Ponovo će se analizirati situacija kada se porede sva tri igrača. Skup podataka će sadržati 139 instanci, koje će predstavljati crno-bele slike putanja zglobova i laka desne ruke. Na slikama ispod, prikazane su po jedna slika za svaku od tri klase:



(a) Primer forhenda Dimitrova



(b) Primer forhenda Sinera



(c) Primer forhenda Alkaraza

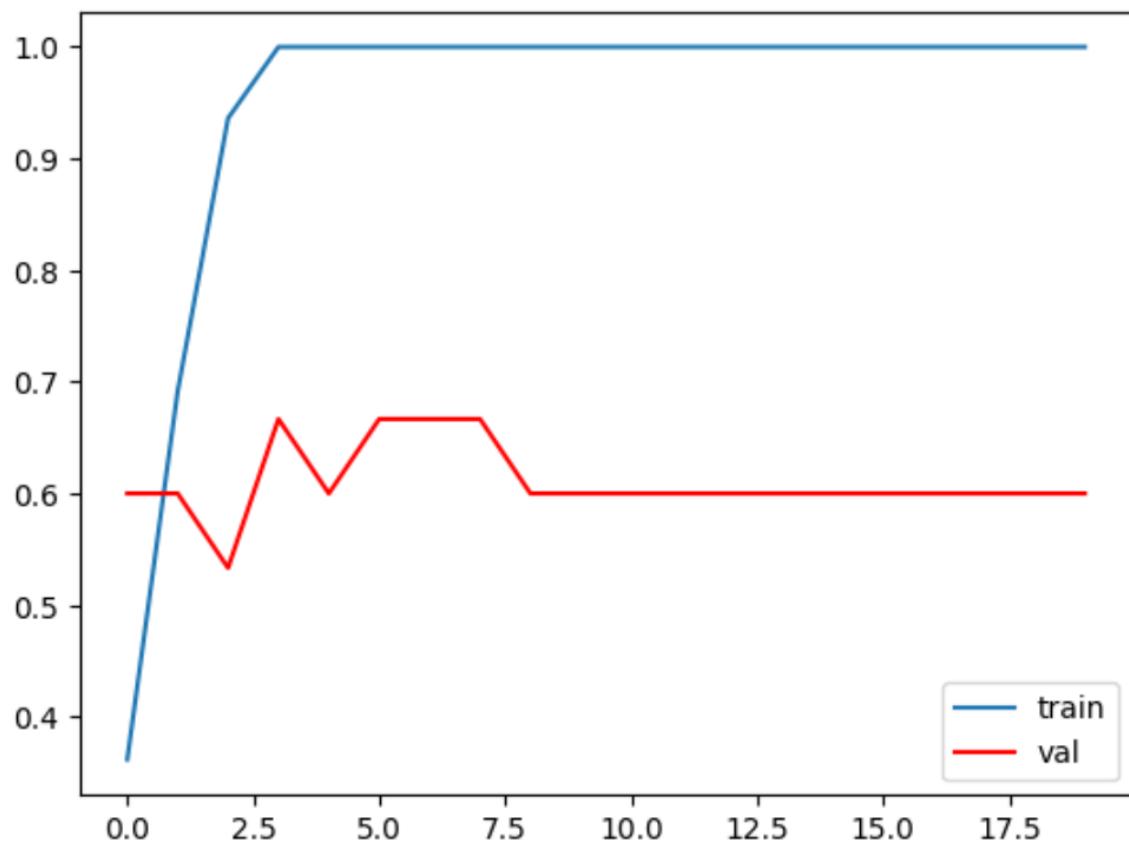
Slika 20: Prikaz instanci skupa

Model će imati sledeću arhitekturu:

```
ConvNet(
    (layer1): Conv2d(3, 16, kernel_size=(3, 3), stride=(1, 1))
    (layer2): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1))
    (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (flatten): Flatten(start_dim=1, end_dim=-1)
    (activation): ReLU()
    (layer3): Linear(in_features=312832, out_features=128, bias=True)
    (fc): Linear(in_features=128, out_features=3, bias=True)
    (dropout): Dropout(p=0.2, inplace=False)
)
```

Slika 21: Arhitektura konvolutivnog modela sa 3 kanala

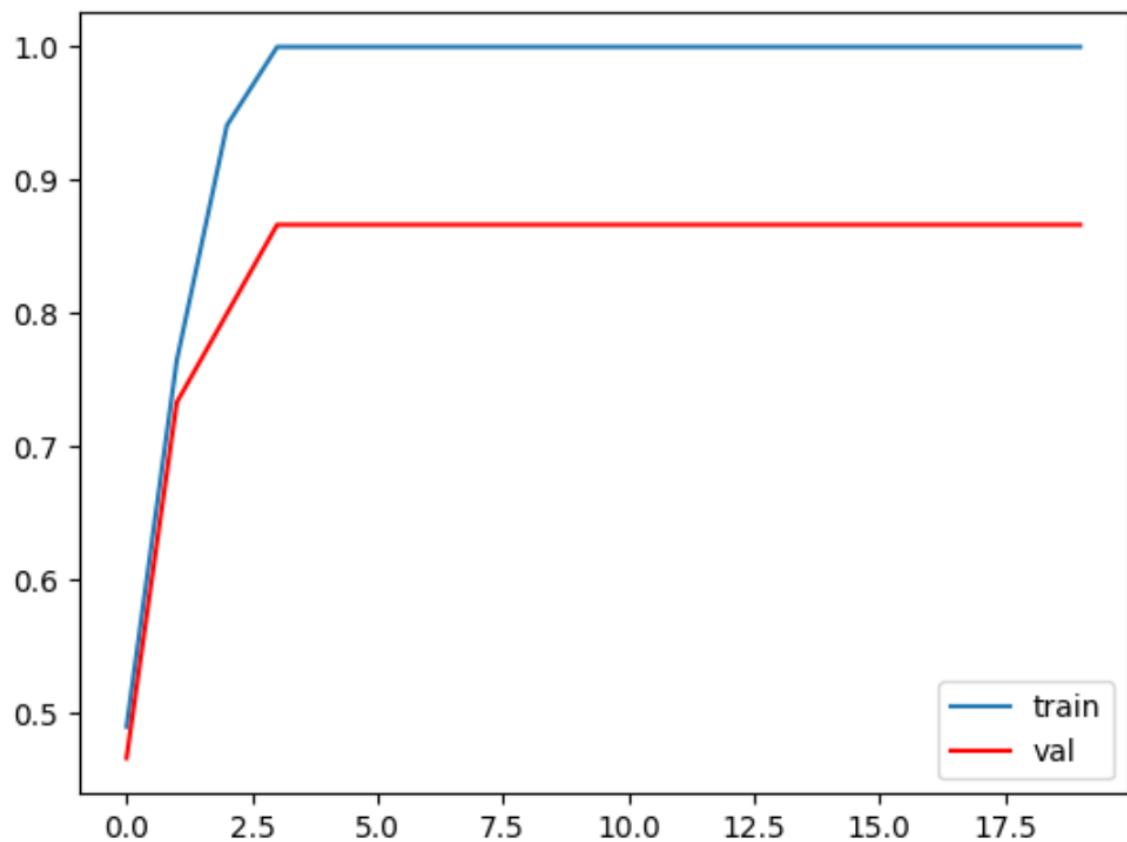
Takođe, i ovaj model će se trenirati 20 epoha. Može se primetiti sa slike da su najbolje karakteristike su ostvarene u 6. epohi:



Slika 22: Tačnost na trening i validacionom skupu kroz epohe

Dobijaju se sledeći rezultati na glavnom test skupu i dodatnom test skupu:
glavni test skup - 80%
dodatni test skup - 62.5%

Kada su u pitanj skupovi Alkaraza i Sinera, najbolji rezultat je u epohi broj 4:



Slika 23: Tačnost na trening i validacionom skupu kroz epohe

Tu se dobijaju sledeći rezultati tačnosti:

glavni test skup - 86.67%

dodatni test skup - 92.86%

4 Zaključak

U ovom radu je pokušana klasifikacija forhenda tenisera. Prvo je bilo potrebno učitati sve zglobove igrača koristeći *Alpha Pose* sistem[3] za detekciju. Sledеći korak je bila normalizacija svih ulaznih podataka koji se mogu dosta razlikovati u zavisnosti od: ugla kamere, pozicije igrača, građe igrača, brzine kretanja itd. Na kraju, za postizanje eksperimentalnih rezultata, korišćena su tri različita modela: potpuno povezana neuronska mreža, konvolutivna neuronska mreža sa 1 kanalom boje i konvolutivna neuronska mreža sa 3 kanala boje. Rezultati analize pokazali su da je moguće uspešno razlikovati pokrete igrača na osnovu njihovih karakterističnih putanja.

Naravno, ovaj projekat ima mnogo potencijala za dalji napredak. Pre svega, proširenje skupa podataka uključivanjem većeg broja igrača može značajno obogatiti trenutnu mrežu i poboljšati rezultate.

Još jedna od tehnika koja je pokušana, ali se nije tako dobro pokazala, jeste perspektivna projekcija. Ona je u nekim situacijama davala odlicne rezultate, dok u drugim je pravila mnogo velike greške. Ova metoda zaslužuje dalju istragu i potencijalno usavršavanje kako bi se postigla doslednija tačnost.

Pored toga, projekat bi se mogao proširiti na način da omogući poređenje dva forhenda i uočavanje razlike između ta dva udarca. Ova funkcionalnost bi bila izuzetno korisna za amatere koji žele da poboljšaju svoju tehniku, jer bi im omogućila detaljniju analizu i identifikaciju oblasti za poboljšanje.

Literatura

- [1] Haotian Zhang, Ye Yuan, Viktor Makoviychuk, Yunrong Guo, Sanja Fidler, Xue Bin Peng, and Kayvon Fatahalian. *Learning Physically Simulated Tennis Skills from Broadcast Videos*. August 2023.
- [2] Lisa Baily, Nghia Truong, Jonathan Lai, and Phong Nguyen. *Stroke Comparison between Professional Tennis Players and Amateur Players using Advanced Computer Vision*. The American School in Japan, Tokyo Techies, Tokyo Coding Club, 2023.
- [3] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. *AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
- [4] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. *RMPE: Regional Multi-person Pose Estimation*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
- [5] Jiefeng Li, Can Wang, Hao Zhu, Yihuan Mao, Hao-Shu Fang, and Cewu Lu. *CrowdPose: Efficient Crowded Scenes Pose Estimation and a New Benchmark*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 10863–10872.