



UNIVERZITET U BANJOJ LUCI
ELEKTROTEHNIČKI FAKULTET



LOKALIZACIJA KLJUČNIH TAČKA TRODIMENZIONALNIH MODELA LJUDSKOG TIJELA

MASTER RAD

Mentor:

prof. dr Vladimir Risojević

Kandidat:

Stefan Novaković

Banja Luka, jul 2024. godine



UNIVERSITY OF BANJA LUKA
FACULTY OF ELECTRICAL ENGINEERING



KEYPOINT LOCALIZATION ON THREE-DIMENSIONAL MODELS OF HUMAN BODY

MASTER'S THESIS

Mentor:

Dr. Vladimir Risojević, assoc. prof.

Candidate:

Stefan Novaković

Banja Luka, July 2024.

Informacije o mentoru i master radu

Mentor:

Dr Vladimir Risojević, vanredni profesor
Univerzitet u Banjoj Luci, Elektrotehnički fakultet

Naslov master rada:

Lokalizacija ključnih tačaka trodimenzionalnih modela ljudskog tijela

Rezime:

Savremeni pristupi rješavanju raznolikih problema koji zahtijevaju analizu trodimenzionalnih (3D) modela i skupova tačaka oslanjaju se na upotrebu algoritama zasnovanih na dubokom učenju koji direktno procesiraju 3D podatke kao što su koordinate tačaka, normalni vektori i informacije o povezanosti površinskih tačaka. U ovom radu predloženo je jedno takvo rješenje za problem pozicioniranja ključnih tačaka tijela i prstiju za animacijske skelete unutar 3D modela ljudskog tijela. S obzirom na oskudicu anotiranih stvarnih 3D skenova ljudi, u ovom radu se generišu sintetički primjeri čiji se parametri oblika i poze variraju. Slično kao i u jednom od savremenih pristupa, predloženi metod izračunava lokaciju svake ključne tačke kao konveksnu kombinaciju ulaznih površinskih tačaka. Za dati ulazni skup površinskih tačaka, upotrebom dinamičke graf-konvolucione neuronske mreže se predviđaju koeficijenti takvih konveksnih kombinacija. Poređenjem predloženog metoda sa pomenutim savremenim pristupom, pokazano je da je moguće postići znatno bolje rezultate sa jednostavnijom arhitekturom, pogotovo za ključne tačke prstiju. S obzirom na to da predloženi algoritam zahtijeva manji broj unaprijed izračunatih obilježja, postižu se i znatno kraća vremena procesiranja.

Ključne riječi: animacijski skelet, trodimenzionalni modeli, estimacija poze, neuronske mreže

Naučna oblast: Inženjerstvo i tehnologija

Naučno polje: Elektrotehnika, elektronika i informacione tehnologije

Klasifikaciona oznaka: T 111

Tip odabrane licence Kreativne zajednice: CC BY-SA

Information about the mentor and the master's thesis

Mentor:

Dr. Vladimir Risojević, assoc. prof.
University of Banja Luka, Faculty of Electrical Engineering

Title of Master's Thesis:

Keypoint Localization on Three-Dimensional Models of Human Body

Abstract:

Contemporary approaches to solving various problems regarding analysis of three-dimensional (3D) meshes and point clouds have adopted the use of deep learning algorithms that directly process 3D data such as point coordinates, normal vectors and vertex connectivity information. This work proposes one such solution to the problem of positioning body and finger animation skeleton joints within 3D models of human bodies. Due to the scarcity of real human 3D scans with appropriate annotations, synthetic samples with varied shape and pose parameters are generated. Similarly to one state-of-the-art approach, the proposed method computes each joint location as a convex combination of input points. Given only a list of input points, a dynamic graph convolutional neural network is used to predict the coefficients of the convex combinations. By comparing the proposed method with the aforementioned state-of-the-art method, it is shown that it is possible to achieve significantly better results with a simpler architecture, especially for finger joints. Since the proposed solution requires fewer precomputed features, it also allows for significantly shorter processing times.

Keywords: animation skeleton, three-dimensional models, pose estimation, neural networks

Scientific area: Engineering and technology

Scientific field: Electrical engineering, electronics and information engineering

Classification code: T 111

Creative Commons license type: CC BY-SA

SADRŽAJ

1	Uvod	1
1.1	Motivacija	1
1.2	Skeleti 3D modela i upotreba skeleta u animaciji	2
1.3	Metode za lokalizaciju ključnih tačaka	3
1.4	Duboko učenje nad trodimenzionalnim modelima	4
1.5	Problem i ciljevi	5
1.6	Doprinosi	6
1.7	Pregled sadržaja	7
2	Pregled literature	8
2.1	Analiza skupova tačaka	8
2.1.1	PointNet	8
2.1.2	PointNet++	11
2.1.3	Dinamička graf-konvoluciona neuronska mreža i EdgeConv modul	12
2.2	Estimacija animacijskog skeleta	14
2.2.1	RigNet	14
2.2.2	TARig	16
2.3	Skupovi podataka i parametarski modeli ljudskog tijela	18
2.3.1	MakeHuman	21
3	Arhitektura rješenja	26
4	Skup podataka i evaluacione metrike	32
4.1	Sintetički podaci	32
4.2	Korekcija dužina kostiju skeleta	33
4.3	Generisanje nasumične poze	34
4.4	Rekonstrukcija topologije i normalizacija koordinata	36
4.5	Estimacija normala	37
4.6	Evaluacione metrike	37

5	Eksperimentalni rezultati	40
5.1	Pregled eksperimenata	40
5.1.1	Obučavanje	41
5.1.2	Evaluacija	44
5.2	Objašnjivost	47
5.3	Analiza grešaka	48
5.4	Vizualna analiza najgorih slučajeva za tačke tijela	50
5.5	Vizualna analiza najgorih slučajeva za tačke prstiju	54
6	Zaključak i budući rad	57

Uz rad je priložen CD.

Glava 1

Uvod

1.1 Motivacija

Digitalni trodimenzionalni modeli ljudskog tijela nalaze primjenu u velikom broju oblasti i koriste se na raznovrsne načine. Neki od domena upotrebe su: animacija, filmska industrija, virtualna i augmentovana realnost, simulacije, zdravstvo i medicina, sport i fitnes, biomehanika, antropometrija, ergonomika, moda i dizajn obuće, interakcija čovjek-računar, biometrija, robotika, autonomna vožnja, sinteza podataka, digitalna umjetnost, trodimenzionalna (3D) štampa, forenzika i arheologija.

3D modeli ljudskog tijela mogu da nastanu na tri relevantna načina. Prvi način je da se ručno izrade pomoću softverskih alata, u procesu 3D modelovanja. Drugi način je da se sintetički generišu, pomoću parametrizovanih 3D modela kod kojih se podešavaju parametri poput visine, mase, proporcija i drugih faktora koji imaju uticaj na oblik i pozu tijela [1]–[8]. Treći način je da nastanu akvizicijom iz stvarnog svijeta, procesom 3D skeniranja [9]–[22]. S obzirom na to da se 3D skeniranjem proizvode digitalne reprezentacije tijela stvarnih osoba, ovaj proces ima ključnu ulogu u prikupljanju stvarnih podataka koji su korisni u industriji i istraživanju.

Relevantne tehnike za 3D skeniranje ljudskog tijela su: fotogrametrija, skeniranje pomoću strukturisanog svjetla, lasersko skeniranje i skeniranje korištenjem *time-of-flight* kamera. Navedene tehnike omogućavaju površinsko skeniranje, na osnovu kojeg se izvode površinski 3D modeli. Takvi 3D modeli modeluju vanjsku površinu tijela, ali ne i unutrašnjost tijela. Površinskim skeniranjem se dobija 3D model u formi skupa tačaka. Procesom rekonstrukcije površine, skup tačaka se može transformisati u poligonsku mrežu, u kojoj su susjedne tačke povezane i formiraju poligone. Pored površinskog skeniranja, postoje i tehnologije za zapreminsko skeniranje tijela ili dijelova tijela, poput kompjuterizovane tomografije i snimanja magnetnom rezonancom, na osnovu kojih se dobija zapreminski 3D model. Takav 3D model reprezentuje unutrašnjost tijela.

U ostatku teksta, pod pojmom 3D modela podrazumijevaju se površinski 3D modeli. Skeniranjem tijela čovjeka na način da se koristi veći broj kamera, postavljenih tako da se osoba snima iz više tački gledišta, moguće je dobiti detaljne digitalne rekonstrukcije tijela osobe.

3D model se može kreirati i kombinacijom više metoda. Na primjer, moguće je da se izvrši skeniranje ljudskog tijela u pozi u kojoj bi se rezultat skeniranja najbolje poklapao sa sintetičkim šablonskim 3D modelom prilagođenih parametara. Nakon toga, moguće je ručno modifikovati 3D model koji je proistekao iz procesa skeniranja. To može biti u svrhu korekcije ili izmjene detalja, kao i u svrhu dodavanja novog sadržaja na 3D model.

U industriji se pojavljuje potreba za preciznom lokalizacijom ključnih tačaka za skenirano tijelo čovjeka, koje mogu biti površinske ključne tačke (*landmarks*) ili ključne tačke skeleta (zglobovi, *joints*). Na osnovu lokacija ključnih tačaka moguće je izvesti zaključke o obilježjima 3D modela, kao što su: orijentacija tijela, lokacije i orijentacije zglobova, udaljenosti između zglobova prema različitim metrikama, dužine kostiju, uglovi između kostiju, proporcije između dijelova tijela, informacije o posturi, aktivnost koja se vrši, zdravstvene i antropometrijske karakteristike, kao i druga obilježja. S obzirom na to, lokalizacija ključnih tačaka je vrlo značajan problem u svim domenima upotrebe 3D modela ljudskog tijela. Ovaj rad se fokusira na ključne tačke skeleta, u kontekstu računarske animacije, o čemu je više riječi dato u nastavku.

1.2 Skeleti 3D modela i upotreba skeleta u animaciji

Skelet 3D modela je kompaktna apstrakcija 3D modela u obliku hijerarhije povezanih ključnih tačaka. Tipovi ključnih tačaka koji se razmatraju zavise od namjene. Tipično, kada se govori o skeletu 3D modela ljudskog tijela, ključne tačke se imenuju prema odgovarajućim dijelovima tijela i pozicioniraju se na mjestima relevantnih zglobova unutar 3D modela. Tako, na primjer, tipovi ključnih tačaka za skelet ljudskog tijela, u kontekstu računarske animacije, često uključuju: tačke kičme, ramena, laktove, ručne zglobove, kukove, koljena, skočne zglobove i središta oba stopala [23]–[25]. Ako se detaljno razmatraju ključne tačke šaka, onda se obuhvataju i ključne tačke prstiju. S obzirom na simetričnost tijela, većina tipova takvih ključnih tačaka ima i lijevu i desnu varijantu.

Uzimajući računarsku animaciju kao primjer domena upotrebe, animiranje 3D modela tijela se tipično vrši pomoću skeleta 3D modela koji se naziva animacijski skelet. U tom kontekstu ključne tačke se nazivaju animacijski zglobovi (*animation joints*). Parovi povezanih ključnih tačaka u hijerarhiji animacijskog skeleta smatraju se kostima. Animacijski skelet se vezuje za 3D model tako što se odrede težine uticaja kostiju skeleta na dijelove modela [26], [27]. Kako bi se animirao 3D model, neophodno je animirati pojedinačne kosti. Ovo se postiže tako što se za svaku kost specificira sekvenca afinih transformacija u vremenu. Takve afine transformacije

se prenose i na dijelove modela na koje kosti imaju uticaj, u skladu sa težinom uticaja. Najčešće se ove transformacije specificiraju kao kombinacije translacija, rotacija i skaliranja. Fizički ubjedljivi pokreti 3D modela tijela mogu da se postignu primarno rotiranjem kostiju, dovodeći model u različite poze.

1.3 Metode za lokalizaciju ključnih tačaka

Lokalizacija ključnih tačaka može da se vrši ručno, ali to predstavlja dugotrajan proces. Za 3D modele ljudskog tijela, danas se često koriste rješenja za automatsku lokalizaciju ključnih tačaka. Ona mogu biti zasnovana na heuristikama, upotrebi parametrizovanih 3D modela i dubokom učenju. U nastavku slijede objašnjenja ovih pristupa.

Heuristički metodi tipično zahtijevaju da je poza tijela unaprijed određena, te se zasnivaju na ručno pisanim algoritmima za pronalaženje ključnih tačaka. Kao primjer, za pronalaženje zglobova ruku moguće je krenuti od tačaka 3D modela koje su po frontalnoj osi tijela najudaljenije od ishodišta. Pretpostavka bi bila da su to tačke koje pripadaju šakama. Analizom susjednih tačaka, uz pretpostavke o proporcijama, mogle bi se procijeniti lokacije ručnih zglobova i laktova. Slično bi se, na primjer, moglo uraditi i za zglobove nogu, počevši od najnižih tačaka. Međutim, u praksi se ovakvi pristupi često zasnivaju na netačnim pretpostavkama, što dovodi do pogrešnih rezultata. S druge strane, heuristički metodi mogu biti korisni za iterativno poboljšavanje inicijalnih procjena koje su nastale pomoću drugih metoda.

Parametrizovani 3D modeli obično uključuju ključne tačke skeleta ili funkciju za izračunavanje lokacija ključnih tačaka skeleta na osnovu postavke parametara oblika [1]–[6]. Optimizacijom parametara moguće je pronaći konfiguraciju kojom se postiže najbolje poklapanje parametrizovanog modela sa 3D skenom. Nakon toga, lokacije ključnih tačaka prilagođenog parametrizovanog modela mogu da se iskoriste kao lokacije ključnih tačaka za 3D sken.

Rješenja zasnovana na dubokom učenju mogu da se direktno obučavaju za lokalizaciju ključnih tačaka 3D modela. 3D duboko učenje može da se vrši nad projekcijama i razlaganjima 3D scena ili objekata. U ovom kontekstu, monokularno razumijevanje podrazumijeva izvođenje zaključaka o 3D svojstvima na osnovu jedne slike ili mape dubine. Moguće je i 3D oblik svesti na 2D oblik korištenjem nelinearnih projekcija. Takođe, snimci iz više perspektiva, dobijeni korištenjem više kamera ili pomjeranjem kamere, mogu se kombinovati. Međutim, ovakvi pristupi tipično imaju problem gubitka informacija o trodimenzionalnim oblicima.

Neuronske mreže mogu i direktno da se obučavaju i izvršavaju nad 3D podacima. Jednostavan format podataka je mreža vokseli, koji se može posmatrati kao trodimenzionalna generalizacija digitalnih slika. Mreža vokseli može nastati u procesu vokselizacije 3D modela koji je inicijalno u formi skupa tačaka ili mreže poligona. U ovom procesu se vrši uniformna diskretizacija zau-

zeća 3D prostora u kojem se model nalazi, čime se konstruiše 3D niz koji predstavlja zauzeće pojedinačnih ćelija u prostoru. Ovakav niz se smatra regularnom mrežom vokselâ. Motivacija za primjenu vokselizacije proistiĉe iz toga što duboko uĉenje nad regularnim mrežama vokselâ zahtijeva jednostavne operacije, poput 3D konvolucije. Praktiĉna upotreba regularnih mreža vokselâ je znaĉajno ograniĉena vremenskom i prostornom sloŹenošću. Mreža vokselâ moŹe biti i rijetko popunjena, tako da se u njoj ĉuvaju samo popunjeni vokseli, što moŹe znatno smanjiti zauzeće memorije i vrijeme obrade. Obrada rijetko popunjenih mreža vokselâ zahtijeva komplikovanije operacije, u poreĉenju sa regularnim mrežama vokselâ. Nedostatak mreža vokselâ je u tome što se procesom vokselizacije mogu izgubiti znaĉajni detalji oblika 3D modela, kao posljedica uniformne diskretizacije. Takoĉe, u zavisnosti od implementacije i Źeljene rezolucije, proces vokselizacije moŹe biti skup po pitanju vremena izvršavanja. Svi ovi faktori daju motivaciju za razvoj neuronskih mreža za direktnu analizu 3D modela, bez potrebe za vokselizacijom. IstraŹivanje u ovom radu ĉe biti fokusirano na metode zasnovane na dubokom uĉenju nad 3D modelima. S obzirom na to, u daljnjem tekstu slijedi opis relevantnih metoda.

1.4 Duboko uĉenje nad trodimenzionalnim modelima

U posljednjih nekoliko godina, razvijene su arhitekture neuronskih mreža koje direktno analiziraju 3D modele u formi skupa taĉaka ili mreŹe poligona. Osnovna prednost direktne obrade 3D modela je u tome što se ne gube detalji oblika usljed diskretizacije ili projekcije. Takoĉe, modeli u ovim formama predstavljaju efikasne reprezentacije, s obzirom na to da im prostorna sloŹenost raste linearno sa brojem taĉaka. Ćak i ako je potrebno da se izvrši pododmjeravanje taĉaka 3D modela, preostale taĉke mogu i dalje da dovoljno detaljno opisuju oblik modela. S druge strane, primarni izazov je u tome što arhitekture neuronskih mreža za direktno procesiranje 3D modela zahtijevaju sloŹene operacije, s obzirom na iregularnu strukturu analiziranih podataka. Neki od problema koji su se morali riješiti kroz takve arhitekture ukljuĉuju: postizanje nezavisnosti izlaza od konkretne permutacije ulaznog skupa taĉaka, odabir metrika za mjerenje udaljenosti izmeĉu taĉaka, definisanje susjedstava izmeĉu taĉaka, ekstrakcija lokalnih obiljeŹja i definisanje operacija konvolucije nad taĉkama.

PointNet [28] je prvobitno predloŹena arhitektura za analizu skupova taĉaka. PointNet++ [29] unapreĉuje PointNet arhitekturu, poboljšavajući mogućnosti ekstrakcije lokalnih obiljeŹja. U [30], predlaŹe se upotreba dinamiĉke graf-konvolucione neuronske mreŹe, zasnovana na tzv. EdgeConv graf-konvolucionom operatoru, ĉime se postiŹu znatna poboljšanja u odnosu na ranije arhitekture. U [31] se predlaŹe PointNeXt arhitektura, izvedena iz PointNet++ arhitekture, a koja nastaje unapreĉenjem strategija obučavanja i augmentacije podataka, kao i uz manje izmjene originalne PointNet++ arhitekture, ĉime se dobijaju znatno bolje performanse na nekoliko standardnih skupova podataka. U [32] se sistematizuju postojeće arhitekture, uvoĉenjem

PointMeta meta-arhitekture za razumijevanje skupova tačaka.

RigNet arhitektura [33] služi za generisanje funkcionalnog animacijskog skeleta za stilizovane likove, bez predispozicije po pitanju tipova ključnih tačaka. U [34] se predlaže TARig arhitektura (*Template-Aware Rig*), koja nastaje adaptacijom RigNet arhitekture za slučaj kada postoje predefinisani tipovi ključnih tačaka. U [35] se rješava problem lokalizacije ključnih tačaka na osnovu skupa 3D tačaka, uz dvofazno predtreniranje zasnovano na sintetičkom skupu podataka i nenadgledanom obučavanju.

Pored ključnih tačaka tijela, može biti neophodno procijeniti i lokacije ključnih tačaka prstiju. Pri procjeni lokacija ključnih tačaka prstiju, tolerancija na greške je tipično manja. Ovo znači da procjene lokacija ključnih tačaka prstiju moraju imati znatno manju apsolutnu udaljenost od stvarnih lokacija, u odnosu na procjene lokacija ključnih tačaka tijela. U literaturi postoje radovi u kojima se vrši estimacija 3D poze šake na osnovu 3D modela šake, zasnovani na PointNet i PointNet++ arhitekturama [36]–[39]. Rješenja predložena u literaturi ne uzimaju u obzir kombinovanu estimaciju 3D ključnih tačaka tijela i prstiju nad 3D modelom ljudskog tijela.

Anotirani skupovi 3D skenova ljudskog tijela, koji bi bili pogodni za obučavanje neuronske mreže za problem lokalizacije ključnih tačaka 3D modela, nisu javno dostupni, ne sadrže dovoljan broj 3D modela ili ne sadrže dovoljno precizne anotacije. Čak i oni skupovi namijenjeni za estimaciju 3D poze ljudskog tijela tipično ne sadrže detaljne 3D modele, s obzirom na to da je skeniranje vršeno sa ograničenim vidnim poljem. Pored toga, akvizicija 3D skenova zahtijeva skupu opremu. Transformacija skenova u upotrebljive 3D modele zahtijeva dodatnu obradu skenova. Takođe, anotiranje 3D skenova je dugotrajan proces. Svi ovi faktori utiču na nedostatak adekvatnih skupova anotiranih 3D skenova. Upotrebom alata za generisanje 3D modela ljudskog tijela moguće je konstruisati sintetički skup podataka [1]–[8]. Takav skup podataka može se koristiti za predtreniranje neuronske mreže, prije finog podešavanja nad stvarnim podacima, kao u [35].

1.5 Problem i ciljevi

Problem koji se u ovom radu razmatra jeste konstrukcija algoritma zasnovanog na dubokom učenju, koji, za predefinisani šablon animacijskog skeleta, na osnovu 3D modela ljudskog tijela procjenjuje pozicije ključnih tačaka šablonskog skeleta. Pri tome, treba biti podržana i precizna lokalizacija ključnih tačaka prstiju. Postojeće neuronske mreže za estimaciju ključnih tačaka 3D modela tijela nisu adekvatne i za estimaciju ključnih tačaka prstiju, čineći taj aspekt nečim što se može unaprijediti. Pri tome, kao evaluacione metrike, korišće se prosječna apsolutna greška lokalizacije ključne tačke, po Euklidovoj udaljenosti, kao i procenat tačno lokalizovanih

ključnih tačaka. Ove evaluacione metrike su preciznije formulisane u Glavi 4.

Umjesto upotrebe stvarnih 3D skenova, formiraće se skup sintetičkih 3D modela ljudskog tijela upotrebom adekvatnih softverskih alata, koji će biti iskorišten u svrhu poređenja algoritama. Projektovano rješenje bi trebalo, u primarnom domenu upotrebe, da smanji količinu vremena neophodnog za kreiranje animacijskog skeleta za 3D modele ljudskog tijela.

U ovom radu, ostvareni su sljedeći ciljevi:

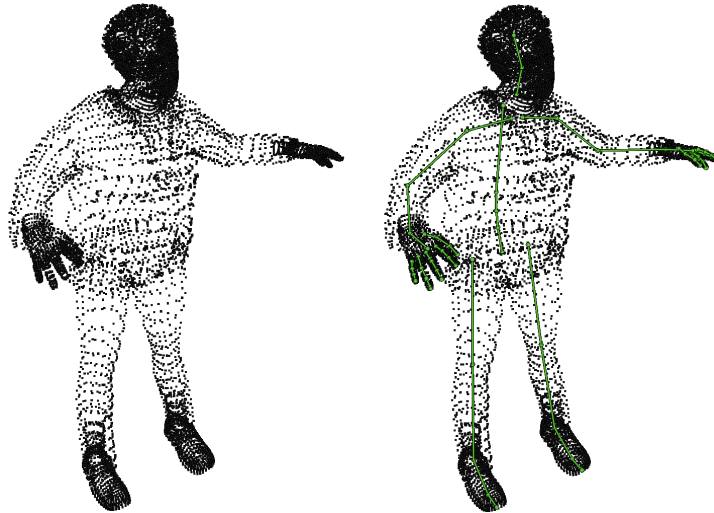
- dati su pregled i analiza savremenih metoda za lokalizaciju ključnih tačaka 3D modela ljudskog tijela,
- predložen je metod za generisanje sintetičkog skupa 3D modela ljudskog tijela sa animacijskim skeletom u varirajućim pozama i generisan je takav skup i
- konstruisana je arhitektura neuronske mreže za lokalizaciju ključnih tačaka 3D modela ljudskog tijela i prstiju, te su obučene i evaluirane konkretne neuronske mreže na osnovu konstruisane arhitekture, uz analizu uticaja nekoliko hiperparametara.

1.6 Doprinosi

Analizom postojeće literature, utvrđene su kandidatske arhitekture koje se mogu prilagoditi problemu lokalizacije ključnih tačaka 3D modela ljudskog tijela. Uz neophodne modifikacije arhitekture i optimizaciju hiperparametara, obučene su i evaluirane neuronske mreže za lokalizaciju ključnih tačaka 3D modela ljudskog tijela, namijenjenih za animacijski skelet. Za evaluaciju, mjerena je prosječna greška, u vidu Euklidove udaljenosti, po tipu ključne tačke, kao i procenat pravilno detektovanih ključnih tačaka, na način opisan u Glavi 4.

Korištenjem postojećih alata za generisanje 3D modela ljudskog tijela sa animacijskim skeletom, generisan je veliki broj primjeraka za obučavanje. Dobijeni skup primjeraka je augmentovan variranjem poze ljudskog tijela i poze šaka, upotrebom animacijskog skeleta.

Kao rezultat istraživanja, predložena je optimizovana arhitektura neuronske mreže za lokalizaciju ključnih tačaka 3D modela ljudskog tijela, koja podržava i preciznu estimaciju ključnih tačaka prstiju. Neuronska mreža koja je obučena na skupu sintetičkih fotorealističnih 3D modela ljudi, sa varijacijama u pozi, obliku i odjeći, proizvodi zadovoljavajuće rezultate u fazi testiranja, prema adekvatnim evaluacionim metrikama. Primjer ulaznog skupa tačaka sa vizuelizovanim izlaznim ključnim tačkama dat je na Slici 1.1. Kao dio istraživačkog rada, objavljen je i naučnoistraživački rad [40] u kojem je prezentovan dio istraživanja u ovom radu.



Slika 1.1: Vizualizacija ulaznog skupa tačaka i istog skupa prekrivenog sa izlaznim procjenama ključnih tačaka, spojenim u skeletalnu hijerarhiju, obojenu u zeleno.

1.7 Pregled sadržaja

U Glavi 2 dat je pregled postojeće literature uz izlaganje neophodnih teorijskih osnova. Obradeni su relevantni metodi, zasnovani na dubokom učenju, za analizu skupova tačaka i za estimaciju animacijskog skeleta. Dat je kratak pregled parametarskih modela ljudskog tijela, kao i detaljan pregled mogućnosti korištenog MakeHuman softverskog alata za generisanje sintetičkih 3D modela ljudskih tijela. U Glavi 3 formulisana je predložena arhitektura neuronske mreže. U Glavi 4 objašnjen je, uz vizuelne primjere, postupak generisanja sintetičkih podataka, kao i njihovog prilagođavanja za korištenje pri obučavanju neuronskih mreža za razmatrani problem lokalizacije ključnih tačaka. Pored toga, formulisane su i korištene evaluacione metrike. U Glavi 5 izloženi su rezultati eksperimenata, uz opis procesa obučavanja i analizu ponašanja neuronske mreže. Predloženo rješenje je upoređeno sa jednim savremenim rješenjem i, prema rezultatima evaluacije, pokazani su bolji rezultati prema dvije korištene metrike. Analizirane su i greške po tipovima ključnih tačaka, uz vizuelne primjere najgorih slučajeva. U Glavi 6 dati su zaključak i prijedlozi za budući rad.

Glava 2

Pregled literature

2.1 Analiza skupova tačaka

2.1.1 PointNet

PointNet [28] je prvi značajan pristup problemu dubokog učenja nad skupovima tačaka. Inicijalno je primijenjen na problemima klasifikacije i segmentacije skupova 3D tačaka. Prije razmatranja doprinosa PointNet pristupa, neophodno je uvesti nekoliko relevantnih definicija, što je urađeno u nastavku.

Teorijske osnove

U nastavku teksta, sa $\mathbb{N}^{\leq n}$ označavaće se skup prirodnih brojeva manjih od ili jednakih n , sa $\mathbb{R}^{\geq 0}$ označavaće se skup nenegativnih realnih brojeva i, za skup S , sa 2^S označavaće se partitivni skup skupa S .

Definicija 1 (Relacija "je permutacija" nad uređenim n -torkama). *Neka su \mathcal{S} neprazan skup, $N \in \mathbb{N}$, $\mathcal{T} = (s_1, s_2, \dots, s_N) \in \bigcup_{K=1}^{\infty} \mathcal{S}^K$ i $\mathcal{B} : \mathbb{N}^{\leq N} \rightarrow \mathbb{N}^{\leq N}$ bijekcija. Tada se kaže da je $(s_{\mathcal{B}(1)}, s_{\mathcal{B}(2)}, \dots, s_{\mathcal{B}(N)})$ permutacija N -torka \mathcal{T} , $\mathcal{T} \stackrel{\pi}{\sim} (s_{\mathcal{B}(1)}, s_{\mathcal{B}(2)}, \dots, s_{\mathcal{B}(N)})$, gdje važi da je $\stackrel{\pi}{\sim} \subseteq (\bigcup_{K=1}^{\infty} \mathcal{S}^K) \times (\bigcup_{K=1}^{\infty} \mathcal{S}^K)$ relacija.*

Može se pokazati da je $\stackrel{\pi}{\sim}$ relacija ekvivalencije. Za $\mathcal{T} \in \bigcup_{K=1}^{\infty} \mathcal{S}^K$, sa $[\mathcal{T}]_{\stackrel{\pi}{\sim}} = \{\tilde{\mathcal{T}} \mid \mathcal{T} \stackrel{\pi}{\sim} \tilde{\mathcal{T}}\}$ označavaće se klasa ekvivalencije kojoj \mathcal{T} pripada. Ona je skup svih permutacija od \mathcal{T} .

Definicija 2 (Simetrična funkcija). *Neka su \mathcal{S} i \mathcal{C} skupovi, $\mathcal{D} \subseteq \bigcup_{N \in \mathbb{N}} \mathcal{S}^N$ i $f : \mathcal{D} \rightarrow \mathcal{C}$. Kaže se da je f simetrična funkcija ako*

$$(\forall \mathcal{T} \in \mathcal{D}) (\forall \tilde{\mathcal{T}} \in [\mathcal{T}]_{\stackrel{\pi}{\sim}}) \tilde{\mathcal{T}} \in \mathcal{D} \wedge f(\tilde{\mathcal{T}}) = f(\mathcal{T}). \quad (2.1)$$

Definicija 3 (Metrički prostor). Neka je \mathcal{M} skup i $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^{\geq 0}$. Kaže se da je (\mathcal{M}, d) metrički prostor ako za svako $x, y, z \in \mathcal{M}$ važi:

1. $d(x, x) = 0$,
2. $x \neq y \implies d(x, y) > 0$ (pozitivnost),
3. $d(x, y) = d(y, x)$ (simetričnost) i
4. $d(x, z) \leq d(x, y) + d(y, z)$ (nejednakost trougla).

U zavisnosti od konteksta, nekada se i sam skup \mathcal{M} naziva metrički prostor, s obzirom na d . Elementi \mathcal{M} se nazivaju tačke metričkog prostora. d se naziva metrika. [41], [42]

Od interesa su, za $N \in \mathbb{N}$, $R \subseteq \mathbb{R}^N$, $R \neq \emptyset$ i metriku d na R , metrički prostori (R, d) .

Definicija 4 (Neprekidnost funkcije između metričkih prostora). Neka su (X, d_X) i (Y, d_Y) metrički prostori. Neka je $f : X \rightarrow Y$, $x_0 \in X$ i neka važi

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists \delta \in \mathbb{R}^+)(\forall x \in X) d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \varepsilon. \quad (2.2)$$

Tada se kaže da je f neprekidna u tački x_0 , s obzirom na metrike d_X, d_Y . Ako za svako $x \in X$ važi da je f neprekidna u tački x , onda se kaže da je f neprekidna funkcija, s obzirom na d_X, d_Y . [41], [42]

Definicija 5 (Hausdorfova udaljenost). Neka je (\mathcal{M}, d) metrički prostor. Neka su $X, Y \subseteq \mathcal{M}$, $X \neq \emptyset$ i $Y \neq \emptyset$. Hausdorfova udaljenost, $d_H : (2^{\mathcal{M}} \setminus \emptyset) \times (2^{\mathcal{M}} \setminus \emptyset) \rightarrow \mathbb{R}$, između skupova X i Y , s obzirom na (\mathcal{M}, d) , se definiše [43], [44] kao

$$d_H(X, Y) = \max \left(\left(\sup_{\mathbf{x} \in X} \inf_{\mathbf{y} \in Y} d(\mathbf{x}, \mathbf{y}) \right), \left(\sup_{\mathbf{y} \in Y} \inf_{\mathbf{x} \in X} d(\mathbf{x}, \mathbf{y}) \right) \right). \quad (2.3)$$

Može se pokazati da za $B \in \mathbb{R}$, $\mathcal{I} = [-B, +B]$, $N \in \mathbb{N}$, $\mathcal{S} = \{s \in 2^{\mathcal{I}^N} \mid |s| \in \mathbb{N}\}$ važi da je (\mathcal{S}, d_H) metrički prostor, tj. da d_H predstavlja metriku kojom se može mjeriti između udaljenost između konačnih skupova tačaka čije su vrijednosti komponentata ograničene [44], [45].

Definicija 6 (Vektorski maksimum). Neka $n, K \in \mathbb{N}$. Tada je $\mathbf{max} : (\mathbb{R}^K)^n \rightarrow \mathbb{R}^K$,

$$\mathbf{max} \left(\begin{pmatrix} \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{K,1} \end{bmatrix}, \begin{bmatrix} x_{1,2} \\ x_{2,2} \\ \vdots \\ x_{K,2} \end{bmatrix}, \dots, \begin{bmatrix} x_{1,n} \\ x_{2,n} \\ \vdots \\ x_{K,n} \end{bmatrix} \end{pmatrix} \right) = \begin{bmatrix} \max_{i \in \mathbb{N} \leq n} x_{1,i} \\ \max_{i \in \mathbb{N} \leq n} x_{2,i} \\ \vdots \\ \max_{i \in \mathbb{N} \leq n} x_{K,i} \end{bmatrix}. \quad (2.4)$$

PointNet arhitektura

Osnovna ideja u PointNet arhitekturi jeste primjena neuronskih pod mreža za nezavisnu ekstrakciju obilježja po tačkama, praćena simetričnom funkcijom (Definicija 2) za agregaciju obilježja. Pomoću takve simetrične funkcije, kao što je agregacija maksimumom po obilježju (Definicija 6), skup lokalnih obilježja tačaka se sažima u globalni vektor obilježja. Na ovaj način postiže se permutacijska invarijantnost, što znači da izmjena redoslijeda ulaznih tačaka ne utiče na izlaz neuronske mreže. Time se postiže i da je i sama neuronska mreža, za izračunavanje određenog izlaznog obilježja, simetrična funkcija ulaznog uređenog skupa tačaka.

Najznačajniji teorijski doprinos PointNet rada je u tome što pokazuje da se funkcije koje preslikavaju skup tačaka u realan broj, sa adekvatnim svojstvom neprekidnosti, mogu aproksimirati do proizvoljne tačnosti, na način da se elementi formule za aproksimaciju mogu implementirati kao neuronske pod mreže koje služe za ekstrakciju obilježja. U nastavku slijedi precizniji opis ovoga, uz neposredan osvrt na to kako se podizrazi u formuli za aproksimaciju mogu posmatrati iz perspektive mašinskog učenja. Neka su:

- $d \in \mathbb{N}$ dimenzionalnost tačaka,
- $n \in \mathbb{N}$ broj tačaka,
- za $m, M \in \mathbb{R}$, $\mathcal{I} = [m, M]$ interval kojem komponente tačaka pripadaju,
- $\mathcal{S} = \{s \in 2^{\mathcal{I}^d} \mid |s| = n\}$ domen skupova od n tačaka i
- $f : \mathcal{S} \rightarrow \mathbb{R}$ funkcija skupa od n tačaka, neprekidna s obzirom na Hausdorfov udaljenost.

U [28] je pokazano da tada postoje:

- $K \in \mathbb{N}$, koji se može posmatrati kao broj globalnih obilježja skupa tačaka,
- $\gamma : \mathbb{R}^K \rightarrow \mathbb{R}$, koja se može implementirati kao ekstraktor izlaznog obilježja iz vektora globalnih obilježja i
- $h : \mathbb{R}^d \rightarrow \mathbb{R}^K$, koja se može implementirati kao dijeljena pod mreža za ekstrakciju obilježja tačaka,

tako da važi da se f može aproksimirati kao

$$f(\{p_1, p_2, \dots, p_n\}) \approx \gamma \circ \mathbf{max}(h(p_1), h(p_2), \dots, h(p_n)). \quad (2.5)$$

Moguća tačnost aproksimacije raste sa porastom K , koje predstavlja dimenzionalnost domena funkcije γ , kao i dimenzionalnost kodomena funkcije h . Pored ovoga, pokazano je i da je izlaz funkcije $\gamma \circ \mathbf{max}$ tada određen sa maksimalno K ulaznih tačaka. Svaka tačka je inicijalno predstavljena svojim prostornim koordinatama. Pored toga, moguće je uz svaku tačku vezati i

dodatna ulazna obilježja, poput vektora normalnog na površinu na kojoj se tačka nalazi. Osnovna PointNet mreža vrši ekstrakciju obilježja tačaka tako što realizuje h kao dijeljenu podmrežu, odnosno kao lanac dijeljenih podmreža i matičnih transformacija, za procesiranje pojedinačnih tačaka. $\max(h(p_1), h(p_2), \dots, h(p_n))$ predstavlja globalni vektor obilježja, koji se može posmatrati kao simetrična funkcija ulaznog niza tačaka. Za problem klasifikacije, γ određuje uvjerenje u pripadnost nekoj od klasa. Konkretno, γ može biti realizovano kao neuronska mreža sa K ulaznih neurona. Za problem segmentacije u y segmenata, izračunati globalni vektor obilježja se, za svaku tačku, konkatenuira sa vektorom skrivenih obilježja te tačke, dužine l , te se dobijena $n \times (l + K)$ matrica, segmentacijskom podmrežom transformiše u $n \times Y$ matricu skorova. Za probleme klasifikacije i segmentacije, zvanična implementacija PointNet arhitekture prilikom obučavanja optimizuje gubitak unakrsne entropije nad izlazom Softmax sloja koji procesira izlaz posljednjeg sloja neuronske mreže.

2.1.2 PointNet++

PointNet arhitektura nije pogodna za probleme koji zahtijevaju detaljno razmatranje lokalnih obilježja koja proističu iz susjedstava tačaka, s obzirom na to da se nikada ne razmatraju udaljenosti između pojedinačnih tačaka. PointNet++ [31] rješava problem ekstrakcije lokalnih obilježja iz susjedstava. Rekurzivno se apstrahuju klasteri tačaka, na način da se čuva samo reprezentativna tačka iz klastera. Uz takvu tačku vezuju se lokalna obilježja čitavog klastera kojem ona pripada. Ovaj princip je konceptualno sličan operaciji konvolucije u konvolucionim neuronskim mrežama. Za svaki klaster tačaka, obilježja se ekstrahuju primjenom PointNet mreže. Ovime se, kroz više stepeni apstrakcije, početni skup tačaka enkoduje u formu više skupova apstraktnih tačaka, sa znatnim brojem obilježja po apstraktnoj tački. Svakom od tih skupova odgovara određena veličina susjedstva, koja postaje sve veća u dubljim slojevima za apstrakciju tačaka. Na prethodno opisan način, formira se multirezoluciona hijerarhija lokalnih obilježja. Klasifikacija, u PointNet++ arhitekturi, vrši se primjenom PointNet mreže nad posljednjim, najmanjim, skupom apstraktnih tačaka. Segmentacija ulaznog skupa tačaka vrši se segmentacijom apstraktnih tačaka, nakon čega se interpolacijom kroz više slojeva određuju klase preostalih tačaka. Ovime se dobija enkoder-dekoder arhitektura sa preskakajućim vezama, nalik U-Net arhitekturama [46]. Međutim, ovakvim pristupom se, pri obučavanju i inferenciji, zanemaruju, kao obilježja, geometrijske relacije između tačaka, kako unutar klastera, tako i između tačaka iz različitih klastera. Slično kao i zvanična implementacija PointNet arhitekture, zvanična PointNet++ implementacija, za probleme klasifikacije i segmentacije, prilikom obučavanja optimizuje gubitak unakrsne entropije nad izlazom Softmax sloja koji procesira izlaz posljednjeg sloja neuronske mreže.

2.1.3 Dinamička graf-konvoluciona neuronska mreža i EdgeConv modul

U [30] razmatra se upotreba dinamičke graf-konvolucione neuronske mreže (DGCNN) za analizu skupova tačaka. Predlaže se korištenje EdgeConv modula, opisanog u nastavku.

EdgeConv modul je vrsta graf-konvolucionog operatora. Za upotrebu u analizi skupova tačaka, zahtijeva da se skup tačaka preformuliše kao graf. Pri tome, tačke su predstavljene čvorovima grafa, dok grane opisuju relacije između tačke i njenih najbližih susjeda. Ako je dat skup tačaka, predstavljenih koordinatama ili vektorima obilježja, algoritmi poput K najbližih susjeda se mogu iskoristiti kako bi se, za svaki čvor, odredili susjedni čvorovi. Po ugledu na PointNet, EdgeConv izračunava svako obilježje tačke primjenom dijeljene podmreže nad obilježjima grana koje se stiču u čvoru koji predstavlja tu tačku, što je praćeno agregacijom maksimum po obilježju. Pri svakom EdgeConv bloku, susjedstva se ponovno izračunavaju na osnovu udaljenosti vektora obilježja tačaka iz prethodnog EdgeConv bloka. Ulančavanjem više EdgeConv blokova u sekvencu omogućava se grupisanje tačaka sa sličnim karakteristikama. Inicijalna susjedstva se mogu, u najjednostavnijem slučaju, odrediti na osnovu Euklidove udaljenosti između tačaka u 3D prostoru. Ako se susjedstva izračunavaju na osnovu udaljenosti vektora obilježja, onda se govori o susjedstvima u semantičkom prostoru. PointNet i PointNet++ se mogu smatrati specijalnim slučajevima DGCNN mreže. Konkretno, PointNet bi bio slučaj da su veličine razmatranih susjedstava 1, tj. da susjedstvo svakog čvora uključuje samo taj isti čvor. PointNet++ bi bio slučaj u kom se fiksni graf particioniše, pri čemu se ne razmatraju udaljenosti u prostoru obilježja.

EdgeConv modul za konvoluciju nad granama grafa

Sljedeći definicija EdgeConv modula. Neka važi sljedeće:

- hiperparametar $f^{\text{in}} \in \mathbb{N}$ predstavlja broj ulaznih obilježja po čvoru,
- hiperparametar $f^{\text{out}} \in \mathbb{N}$ predstavlja broj izlaznih obilježja po čvoru,
- $\mathcal{M} : \mathbb{N}^{\leq f^{\text{out}}} \rightarrow (\mathbb{R}^{2f^{\text{in}}} \rightarrow \mathbb{R})$ je konačan niz od f^{out} ekstraktora obilježja grana, t.d. $(\forall k \in \mathbb{N}^{\leq f^{\text{out}}}) \mathcal{M}_k : \mathbb{R}^{2f^{\text{in}}} \rightarrow \mathbb{R}$ predstavlja ekstraktor k -tog izlaznog obilježja grane,
- $\mathcal{A} : \bigcup_{N=1}^{\infty} \mathbb{R}^N \rightarrow \mathbb{R}$ je simetrična funkcija koja predstavlja agregator obilježja,
- $\mathcal{V} = \{1, 2, \dots, |\mathcal{V}|\}$ predstavlja skup čvorova grafa,
- ulaz $\mathcal{N} : \mathcal{V} \rightarrow 2^{\mathcal{V}}$, tako da $\mathcal{N}_i \in 2^{\mathcal{V}} \setminus \emptyset$ predstavlja skup susjeda čvora $i \in \mathcal{V}$ i
- ulaz $\mathbf{F}^{\text{in}} \in \mathbb{R}^{|\mathcal{V}| \times f^{\text{in}}}$, tako da $\mathbf{F}_{i,\cdot}^{\text{in}} \in \mathbb{R}^{f^{\text{in}}}$ predstavlja ulazni vektor obilježja za čvor $i \in \mathcal{V}$.

Neka su $i \in \mathcal{V}$ i $j \in \mathcal{N}_i$. Vektor ulaznih obilježja za granu (i, j) je definisan kao konkatencija

vektora obilježja razmatranog čvora i i razlike između tog vektora i vektora obilježja susjeda j ,

$$\mathbf{e}_{i,j} = \begin{bmatrix} \mathbf{F}_i^{\text{in}} \\ \mathbf{F}_j^{\text{in}} - \mathbf{F}_i^{\text{in}} \end{bmatrix} \in \mathbb{R}^{2f^{\text{in}}}. \quad (2.6)$$

Neka je $k \in \{1, 2, \dots, f^{\text{out}}\}$. Tada se k -to izlazno obilježje grane (i, j) izračunava kao

$$e'_{i,j,k} = \mathcal{M}_k(\mathbf{e}_{i,j}) \in \mathbb{R}. \quad (2.7)$$

Za čvor i , k -to izlazno obilježje se izračunava agregacijom obilježja grana incidentnih čvoru i ,

$$f_{i,k}^{\text{out}} = \mathcal{A}(e'_{i,1,k}, e'_{i,2,k}, \dots, e'_{i,|\mathcal{N}_i|,k}) \in \mathbb{R}. \quad (2.8)$$

Vektor izlaznih obilježja za čvor i je predstavljen sa

$$\mathbf{f}_i^{\text{out}} = \begin{bmatrix} f_{i,1}^{\text{out}} & f_{i,2}^{\text{out}} & \dots & f_{i,f^{\text{out}}}^{\text{out}} \end{bmatrix}^T \in \mathbb{R}^{f^{\text{out}}}. \quad (2.9)$$

Izlaz EdgeConv bloka je matrica izlaznih obilježja čvorova,

$$\text{EdgeConv}_{f^{\text{in}}, f^{\text{out}}}(\mathcal{N}, \mathbf{F}^{\text{in}}, \mathcal{M}, \mathcal{A}) = \begin{bmatrix} \mathbf{f}_1^{\text{out}} & \mathbf{f}_2^{\text{out}} & \dots & \mathbf{f}_{|\mathcal{V}|}^{\text{out}} \end{bmatrix}^T \in \mathbb{R}^{|\mathcal{V}| \times f^{\text{out}}}. \quad (2.10)$$

Za ekstraktor obilježja, \mathcal{M} , često se koristi, kao dijeljena podmreža, potpuno povezana klasična neuronska mreža (*fully-connected feed-forward neural network*). Takva mreža se u literaturi često naziva, pomalo neprecizno, višeslojni perceptron (*multi-layer perceptron*, MLP). Ona se može implementirati pomoću konvolucionog operatora s kernelom reda 1 i korakom 1, koji ima onoliko ulaznih kanala koliko ima ulaznih obilježja i onoliko izlaznih kanala (mapa obilježja) koliko ima izlaznih obilježja. U tom kontekstu, \mathcal{M}_k se može smatrati $\mathbb{R}^{2f^{\text{in}}} \rightarrow \mathbb{R}$ funkcijom čije je pravilo preslikavanja predstavljeno formulom za izračunavanje k -tog izlaza neuronske mreže. Kao agregator obilježja, \mathcal{A} , u originalnom radu se koristi max funkcija.

Glavni dio DGCNN mreže je sekvenca EdgeConv modula. Za prvi EdgeConv modul, ulazna obilježja mogu da uključuju, u najjednostavnijem slučaju, 3D koordinate ulaznih tačaka. Takođe, mogu i da uključuju, za svaku tačku, i druga unaprijed izračunata obilježja tačke, poput vektora normalnog na dio površine na kom se tačka nalazi. Naknadni EdgeConv moduli procesiraju, kao ulazna obilježja tačaka, izlazna obilježja iz prethodnih EdgeConv modula. Za probleme klasifikacije i segmentacije, obilježja ekstrahovana iz svih EdgeConv modula se concateniraju i procesiraju podmrežom nad čijim izlazom se izračunava vektor maksimuma po obilježju kako bi se dobio globalni vektor obilježja. Slično PointNet arhitekturi, globalni vektor obilježja se može procesirati klasičnom klasifikacionom podmrežom, kako bi se izvršila klasi-

fikacija ulaznog skupa tačaka, a može se i konkatenerirati sa skrivenim obilježjima iz EdgeConv slojeva, što onda služi kao ulaz za klasičnu segmentacijsku podmrežu. Slično kao i zvanične implementacija PointNet i PointNet++ arhitektura, zvanična implementacija DGCNN arhitekture, za probleme klasifikacije i segmentacije, prilikom obučavanja optimizuje gubitak unakrsne entropije nad izlazom Softmax sloja koji procesira izlaz posljednjeg sloja neuronske mreže.

Dinamičko preračunavanje grafa

U originalnoj arhitekturi, prije obrade skupa tačaka u EdgeConv bloku, susjedstva tačaka se izračunavaju na osnovu udaljenosti u prostoru obilježja. Prema tome, mreža ne obrađuje fiksni graf u svakom koraku, s obzirom na to da se struktura grafa mijenja, odnosno, skup grana grafa se više puta izračunava u toku obrade, na osnovu odgovarajuće matrice obilježja. Ovime je omogućeno da receptivno polje za svaku tačku može da obuhvati bilo koju drugu tačku iz ulaznog skupa tačaka. Konkretno, u originalnoj implementaciji, izračunava se matrica udaljenosti po parovima tačaka, u prostoru obilježja, te se za svaku tačku uzima K najbližih tačaka za formiranje susjedstva te tačke. Susjedstvo izračunato na osnovu matrice obilježja \mathbf{F} će se, u nastavku, označavati sa $\mathcal{N}^{\mathbf{F}}$.

2.2 Estimacija animacijskog skeleta

2.2.1 RigNet

RigNet arhitektura [33], namijenjena za generisanje animacijskog skeleta na osnovu poligonskog 3D modela, sadrži podmrežu za estimaciju pozicija zglobova animacijskog skeleta, koja je zasnovana na kombinaciji EdgeConv modula, modula za učenje pomjeraja tačaka [47] i mehanizma pažnje. U navedenoj arhitekturi, poligonski 3D model treba da se preformuliše kao graf, tako da površinske tačke 3D modela postanu čvorovi grafa. Izdvajanje obilježja se vrši formiranjem skupa grana za topološke susjede, kao i skupa grana za geodetske susjede. Osnovna ideja jeste da mreža nauči izmjestiti površinske tačke ka lokacijama njima najbližih animacijskih zglobova, nakon čega se dobijeni klasteri tačaka grupišu u jedinstvene tačke [48]. Iterativnim ponavljanjem ove operacije, posljednji klasteri bi trebalo da se nalaze na pozicijama ključnih tačaka. Pri tome, u konkretnoj arhitekturi, količina lokalizovanih ključnih tačaka zavisi od konkretnog 3D modela, kao i od parametara algoritma za kontrakciju i klasterizaciju tačaka.

Za određivanje geodetskih susjeda, njihov metod zahtijeva izračunavanje površinskih geodetskih distanci između površinskih tačaka, koje oni implementiraju kao operaciju kompleksnosti $\Theta(n^2 \log n)$, s obzirom na to da, za svaku tačku, sortiraju udaljenosti te tačke do svih drugih tačaka. Ovo ima znatan uticaj na vrijeme potrebno da se procesira 3D model.

Njihov metod nije prikladan za lokalizaciju ključnih tačaka predefinisanih tipova, tj. koji odgovaraju nekom šablonu skeleta, jer se njime samo procjenjuju pozicije na kojima ključne tačke treba da budu, bez uzimanja u obzir njihovih tipova.

GMEdgeConv operator

U RigNet arhitekturi, uveden je GMEdgeConv operator kao osnovni gradivni blok mreže, zasnovan na EdgeConv operatoru. Slijedi definicija GMEdgeConv operatora. Neka važi sljedeće:

- hiperparametar $f^{\text{in}} \in \mathbb{N}$ predstavlja broj ulaznih obilježja po čvoru,
- hiperparametar $f^{\text{out}} \in \mathbb{N}$ predstavlja broj izlaznih obilježja po čvoru, $f^{\text{out}} \bmod 2 \equiv 0$,
- $\mathcal{V} = \{1, 2, \dots, |\mathcal{V}|\}$ predstavlja skup čvorova grafa,
- ulaz $\mathcal{N}^t : \mathcal{V} \rightarrow 2^{\mathcal{V}}$, tako da $\mathcal{N}_i^t \subseteq \mathcal{V}$ predstavlja skup topoloških susjeda čvora $i \in \mathcal{V}$,
- ulaz $\mathcal{N}^g : \mathcal{V} \rightarrow 2^{\mathcal{V}}$, tako da $\mathcal{N}_i^g \subseteq \mathcal{V}$ predstavlja skup geodetskih susjeda čvora $i \in \mathcal{V}$,
- ulaz $\mathbf{F}^{\text{in}} \in \mathbb{R}^{|\mathcal{V}| \times f^{\text{in}}}$, tako da $\mathbf{F}_{i,\cdot}^{\text{in}} \in \mathbb{R}^{f^{\text{in}}}$ predstavlja ulazni vektor obilježja za čvor $i \in \mathcal{V}$,
- $\text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^t$ predstavlja mrežu za ekstrakciju obilježja grana između topoloških susjeda,
- $\text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^g$ predstavlja mrežu za ekstrakciju obilježja grana između geodetskih susjeda,
- $\mathbf{E}^t = \text{EdgeConv}_{f^{\text{in}}, \frac{1}{2}f^{\text{out}}}(\mathcal{N}^t, \mathbf{F}^{\text{in}}; \text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^t, \text{max})$, prema (2.10),
- $\mathbf{E}^g = \text{EdgeConv}_{f^{\text{in}}, \frac{1}{2}f^{\text{out}}}(\mathcal{N}^g, \mathbf{F}^{\text{in}}; \text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^g, \text{max})$, prema (2.10) i
- $\mathcal{M} = (\text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^t, \text{MLP}_{f^{\text{in}}, f^{\text{out}}/2}^g, \text{MLP}_{f^{\text{out}}, f^{\text{out}}})$ predstavlja ekstraktore obilježja.

GMEdgeConv operator se onda definiše kao

$$\text{GMEdgeConv}_{f^{\text{in}}, f^{\text{out}}}(\mathcal{N}^t, \mathcal{N}^g, \mathbf{F}^{\text{in}}; \mathcal{M}) = \text{MLP}_{f^{\text{out}}, f^{\text{out}}}\left(\left[\mathbf{E}^t \mid \mathbf{E}^g\right]_{|\mathcal{V}| \times f^{\text{out}}}\right) \in \mathbb{R}^{|\mathcal{V}| \times f^{\text{out}}}. \quad (2.11)$$

Prilikom obučavanja modula za zglobove, za svaku tačku se uče vektorski pomjeraji ka najbližim ključnim tačkama. Nakon procesiranja ulaznog skupa tačaka sa nizom GMEdgeConv slojeva i dijeljenom podmrežom, procijenjeni pomjeraji se sabiraju sa originalnim tačkama, čime bi trebalo da se dobiju klasteri tačaka. Ideja je da se ovi klasteri nalaze na lokacijama ključnih tačaka. Kao termin ukupne greške, ovdje se računa normalizovana simetrična čamfer udaljenosti između skupa stvarnih ključnih tačaka i skupa pomjerenih tačaka. Zatim, u više ponovljenih koraka, pomjerene tačke prolaze kroz proces kontrakcije i klasterizacije, te se za svaki ovaj korak u ukupnu grešku računa normalizovana simetrična čamfer udaljenost između skupa stvarnih ključnih tačaka i skupa tačaka dobijenog nakon klasterizacije u tom koraku, podijeljena sa ukupnim brojem koraka u kojima se vrši kontrakcija i klasterizacija. Modul za zglobove, dakle, uči da minimizuje ovu sumu normalizovanih simetričnih čamfer udaljenosti.

Evaluacione metrike

U [33] se predlažu i koriste, pored ostalih, dvije relevantne metrike za evaluaciju lokalizovanih animacijskih zglobova: CD-J2J (*Chamfer Distance - Joint to Joint*) i IoU (*Intersection Over Union*).

Sa CD-J2J se mjeri prosječna simetrična normalizovana čamfer udaljenost (Definicija 7), s obzirom na Euklidovu udaljenost, između skupa stvarnih zglobova i skupa predviđenih zglobova.

Definicija 7 (Prosječna simetrična normalizovana čamfer udaljenost). *Neka važi da su $N \in \mathbb{N}$, $X, Y \subset \mathbb{R}^N$, $|X|, |Y| \in \mathbb{N}$ i $d : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ metrika. Tada je prosječna simetrična normalizovana čamfer udaljenost između X i Y , s obzirom na d ,*

$$d_C = \left(\frac{1}{2|X|} \sum_{x \in X} \min_{y \in Y} d(\mathbf{x}, \mathbf{y}) \right) + \left(\frac{1}{2|Y|} \sum_{y \in Y} \min_{x \in X} d(\mathbf{x}, \mathbf{y}) \right). \quad (2.12)$$

Sa IoU metrikom se mjeri procenat tačno lokalizovanih zglobova. U tu svrhu se izračunava, za svaki zglob, lokalni prečnik oblika [49]. Kao prvi korak izračunavanja, vrši se emitovanje zraka, s ishodištem u lokaciji stvarnog zgloba, okomito na kosti spojene tim zglobom. Zatim se pronalaze lokacije na površini 3D modela pogođene emitovanim zrakama, te se izračunava prosječna udaljenost od stvarnog zgloba do njih. Kao lokalni prečnik oblika uzima se dvostruka vrijednost izračunate prosječne udaljenosti. Kako bi se odredio prag udaljenosti za prihvatanje lokalizovanog zgloba kao tačnog, lokalni prečnik oblika se množi sa faktorom u intervalu $[0, 1]$. Mjerenjem broja tačno lokalizovanih zglobova u zavisnosti od faktora praga udaljenosti, može se formirati IoU kriva. Kao motivaciju za ovakav pristup određivanju praga udaljenosti za prihvatanje lokalizovanih zglobova, autori navode, kao primjer, da tanji dijelovi tijela, npr. ruke, imaju manji lokalni prečnik oblika, te, kao rezultat toga, male devijacije mogu uzrokovati primjetna odstupanja u odnosu na deblje dijelove, poput torza. Kao informativnu skalarnu vrijednost za ovu metriku, koriste pragove izračunate sa fiksiranim faktorom 0,5 (IoU@0.5), što znači da izračunati pragovi mogu da se smatraju lokalnim poluprečnicima oblika.

2.2.2 TARig

U TARig arhitekturi [34], modul za šablonske zglobove nastaje kao modifikacija RigNet modula za predikciju zglobova i omogućava lokalizaciju šablonskih zglobova. Pri tome, za ovu svrhu, odbacuje upotrebu operacije za klasterisanje tačaka, te se oslanja na mogućnost predstavljanja lokacija šablonskih zglobova kao konveksnih kombinacija površinskih tačaka 3D modela.

Osnovna mreža za ekstrakciju obilježja tačaka

Neka je $\mathcal{V} = \{1, 2, \dots, |\mathcal{V}|\}$ skup čvorova grafa. Neka su $\mathcal{N}^t, \mathcal{N}^g \in \mathcal{V} \rightarrow 2^{\mathcal{V}}$ geodetska i topološka susjedstva, respektivno. Neka je $\mathbf{F}^{\text{in}} \in \mathbb{R}^{|\mathcal{V}| \times 6}$ matrica ulaznih obilježja, koja su, za svaku tačku, njene koordinate i normalni vektor. Prvo se, pomoću GMEdgeConv slojeva, ekstrahuju obilježja tačaka:

$$\mathbf{F}_1 = \text{GMEdgeConv}_{6,64}(\mathcal{N}^t, \mathcal{N}^g, \mathbf{F}^{\text{in}}; \mathcal{M}_1) \in \mathbb{R}^{|\mathcal{V}| \times 64}, \quad (2.13)$$

$$\mathbf{F}_2 = \text{GMEdgeConv}_{64,256}(\mathcal{N}^t, \mathcal{N}^g, \mathbf{F}_1; \mathcal{M}_2) \in \mathbb{R}^{|\mathcal{V}| \times 256} \text{ i} \quad (2.14)$$

$$\mathbf{F}_3 = \text{GMEdgeConv}_{256,512}(\mathcal{N}^t, \mathcal{N}^g, \mathbf{F}_2; \mathcal{M}_3) \in \mathbb{R}^{|\mathcal{V}| \times 512}, \quad (2.15)$$

gdje su sa $\mathcal{M}_1, \mathcal{M}_2$ i \mathcal{M}_3 predstavljeni obučeni ekstraktori obilježja, prema definiciji GMEdgeConv modula (2.11).

Obilježja iz GMEdgeConv slojeva se spajaju i procesiraju obučenom dijeljenom podmrežom,

$$\mathbf{F}_4 = \text{MLP}_{(64+256+512),1024} \left(\left[\mathbf{F}_1 \mid \mathbf{F}_2 \mid \mathbf{F}_3 \right]_{|\mathcal{V}| \times (64+256+512)} \right) \in \mathbb{R}^{|\mathcal{V}| \times 1024}. \quad (2.16)$$

Agregacijom maksimum po obilježju konstruiše se globalni vektor obilježja,

$$\mathbf{F}_{\text{global}} = \left[\max \mathbf{F}_{4, \cdot, 1} \quad \max \mathbf{F}_{4, \cdot, 2} \quad \dots \quad \max \mathbf{F}_{4, \cdot, |\mathcal{V}|} \right] \in \mathbb{R}^{1 \times 1024}. \quad (2.17)$$

Spajanjem obilježja, uz ponavljanje globalnih obilježja, konstruiše se konačna matrica obilježja,

$$\mathbf{F} = \left[\begin{array}{c|c|c|c|c} \mathbf{F}^{\text{in}} & \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 & \begin{bmatrix} \mathbf{F}_{\text{global}} \\ \mathbf{F}_{\text{global}} \\ \vdots \\ \mathbf{F}_{\text{global}} \end{bmatrix}_{|\mathcal{V}| \times 1024} \end{array} \right] \in \mathbb{R}^{|\mathcal{V}| \times (6+64+256+512+1024)}. \quad (2.18)$$

Modul za šablonske zglobove

Neka je J broj zglobova. \mathbf{F} se transformiše u $|\mathcal{V}| \times J$ matricu elemenata u intervalu $(0, 1)$,

$$\mathbf{H} = \text{Sigmoid}(\text{MLP}_{(6+64+256+512+1024),J}(\mathbf{F})) \in (0, 1)^{|\mathcal{V}| \times J}. \quad (2.19)$$

Za svaki zglob $j \in \mathbb{N}^{\leq J}$, izračunava se, normalizacijom sigmoida, pažnja prema čvoru $i \in \mathcal{V}$,

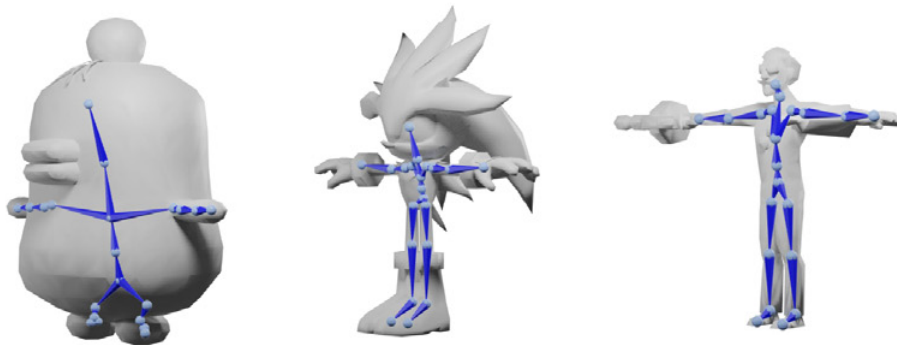
$$a_{ij} = \frac{\mathbf{H}_{ij}}{\sum_{k=1}^{|\mathcal{V}|} \mathbf{H}_{kj}} \in \mathbb{R}. \quad (2.20)$$

Ako je \mathbf{p}_i pozicija i -te površinske tačke, onda se pozicija j -tog zgloba izračunava kao

$$\mathbf{J}_j = \sum_{i=1}^{|\mathcal{V}|} a_{ij} \mathbf{p}_i \in \mathbb{R}^3. \quad (2.21)$$

Prilikom obučavanja optimizuje se suma *dice* gubitka [50] između predikcija mapa pažnje (a_{ij}) i, na način opisan u [34], unaprijed izračunatih mapa pažnje, prosječne kvadratne greške (MSE) između stvarnih i estimiranih lokacija zglobova i normalizovane simetrične čamfer udaljenosti između skupa stvarnih ključnih tačaka i skupa procijenjenih lokacija ključnih tačaka.

U [34] je korišteno 1.636 modela za obučavanje, 205 modela za validaciju i 205 modela za testiranje. Mreža je obučavana 100 epoha, sa *batch* veličinom 2, BatchNorm normalizacijom, Adam optimizatorom [51] i inicijalnim koeficijentom učenja $5e-4$ koji se u pedesetoj epohi smanjuje na $2,5e-4$. Primjeri predikcija šablonskih zglobova iz TARig rada, nad skupom testnih podataka koji je korišten u njihovom radu, dat je na Slici 2.1.



Slika 2.1: Primjer predikcija TARig modula za šablonske zglobove, nad testnim podacima korištenim u njihovom radu, preuzet iz samog rada [34]. Na ovim primjerima je vidljivo da skup podataka koji je korišten u njihovom radu sadrži nerealistične, stilizovane, likove, te da se ne vrši lokalizacija zglobova prstiju.

2.3 Skupovi podataka i parametarski modeli ljudskog tijela

CAESAR [9] (*Civilian American and European Surface Anthropometry Resource Project*) je prvi značajniji komercijalni projekat 3D skeniranja velikog broja ljudi. Rezultovao je skupom od 2.400 skenova građana iz SAD i Kanade, kao i 2.000 skenova građana Evrope. Ovaj skup nije javno dostupan i ne sadrži ključne tačke za animacijski skelet. S obzirom na cilj projekta da se izvrše antropometrijska mjerenja, sve skenirane osobe su minimalno odjevene, te su pozicionirane u A pozi.

Human3.6M [10] je skup koji se sastoji od 3.6 miliona slika sa anotacijama u vidu 3D poza osoba. Skup je kreiran snimanjem, uz praćenje pokreta i primjenu tehnika 3D skeniranja, 11

profesionalnih glumaca u 17 scenarija svakodnevnih aktivnosti. Moguće je i rekonstruisati 3D modele na osnovu ovih podataka. Međutim, s obzirom na to da je skenirano samo 11 različitih osoba, ne postoje značajne varijacije u oblicima tijela. Takođe, stvarne 3D poze s kojima su slike anotirane ne uključuju detaljne poze šaka, tj. ne uključuju ključne tačke prstiju.

FAUST [11] (*Fine Alignment Using Scan Texture*) skup podataka sadrži 300 skenova visoke rezolucije dobijenih skeniranjem 10 različitih osoba u 30 različitih poza. Međutim, ne sadrži anotacije za ključne tačke skeleta 3D modela. Dyna [12] skup podataka se sastoji od 40.000 skenova 10 minimalno odjevenih ljudi u pokretu. Skup ne sadrži anotacije za ključne tačke skeleta. BUFF [13] skup podataka se sastoji od 13.600 skenova, dobijenih skeniranjem 6 osoba u dva različita stila oblačenja i tri različite sekvence pokreta. Ni ovaj skup podataka ne sadrži anotacije za ključne tačke skeleta 3D modela. 3DBodyTex.v2 [14] skup podataka se sastoji od 3.000 3D skenova ljudskih tijela sa teksturama visoke rezolucije. Pri tome, skenirano je 500 osoba, uz značajne raznolikosti u pozama i stilovima odjeće. Međutim, sam skup podataka ne sadrži anotacije za ključne tačke skeleta 3D modela. NOMO3D [15] skup podataka se sastoji od 375 3D skenova različitih osoba u A pozici. Namijenjen je za antropometrijska mjerenja i ne sadrži anotacije za ključne tačke skeleta. U [16], kreirani su BODY-fit i BODY-rgb skupovi podataka koji se sastoje 4.149 i 194 3D skenova, respektivno. Namijenjeni su za antropometrijska mjerenja i ne sadrže anotacije za ključne tačke skeleta. THuman1.0 [17] je skup podataka koji se sastoji od 7.000 nisko-rezolucionih skenova 235 različitih osoba u različitim pozama i u različitoj odjeći, ali ne sadrži anotacije za ključne tačke skeleta. SIZER [18] skup podataka sadrži 2.000 3D skenova u A pozici, nastalih skeniranjem 100 osoba sa varijacijama u odjeći i veličini odjeće. Sam skup podataka ne sadrži anotacije za ključne tačke skeleta 3D modela. THuman 2.0 [19] je skup podataka koji se sastoji od 500 visokorezolucionih skenova 500 osoba. THuman 2.1 proširuje broj modela na 2.500. Skenirane osobe su u različitim pozama i u različitoj odjeći. Ovaj skup podataka ne sadrži anotacije za ključne tačke skeleta. SHARP (Shape-Aware Reconstruction of People) rad [20] uvodi 3DHumans skup podataka koji se sastoji od 180 3D modela nastalih skeniranjem osoba u različitim pozama i sa raznovrsnim stilovima oblačenja. MVP-Human (*Multi-View and Multi-Pose 3D Human*) [21] skup podataka se sastoji od 6.000 skenova 400 različitih osoba. Pri tome, svaka osoba je skenirana, u različitim pozama, ukupno 15 puta. Ovaj skup podataka sadrži anotacije za lokacije ključnih tačaka skeleta, ali ne i za zglobove prstiju. THuman3.0 [22] skup podataka se sastoji od skenova osoba u različitoj odjeći i u različitim pozama. Sadrži 154 kombinacije osobe i skupa odjevnih predmeta, pri čemu je, za svaku kombinaciju, osoba skenirana u 30 do 60 različitih poza. Ovaj skup podataka ne sadrži anotacije za ključne tačke skeleta.

A Skinned Multi-Person Linear Model (SMPL) [1] je naučen statistički model oblika ljudskog tijela koji uzima u obzir i varijacije oblika usljed deformacije poze. Kako bi se modelirale varijacije u obliku, koristi se analiza principijelnih komponentni (PCA), tako da se kao parametri,

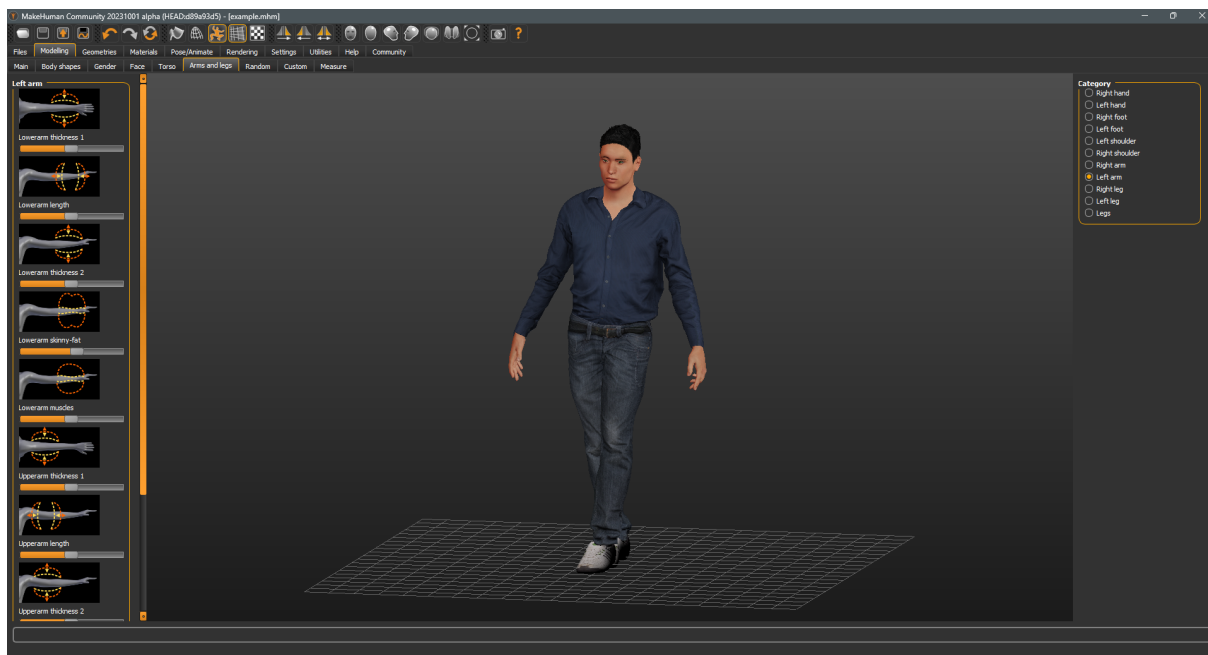
u odgovarajućem redoslijedu, izdvajaju najuticajnije komponente oblika. Na primjer, nekoliko prvih komponenti kontrolišu visinu, debljinu i proporcije između dijelova tijela. U originalnom radu, model je obučavan nad CAESAR [9] skupom podataka.

SMPL+H [2] proširuje SMPL model tako da se uključuju i detaljni pokreti šaka i prstiju. SMPL-X [3], kao parametarski model, proširuje SMPL tako da se uključuju i detaljni pokreti šaka i prstiju, kao i ekspresije lica. GHUM [5] (*Generative Human Unified Model*) je generativni model, zasnovan na upotrebi varijacionog autoenkodera koji uči latentni prostor oblika i poza ljudskog tijela. GHUML [5] (*Generative Human Unified Model with Locomotion*) predstavlja varijantu GHUM modela koja uzima u obzir i pokrete tijela, a time i interakcije između dijelova tijela koje nastaju tokom kretanja. STAR [4] (*A Sparse Trained Articulated Human Body Regressor*) predstavlja poboljšanje u odnosu na SMPL model, uz značajnu redukciju broja parametara i bolju generalizaciju. SUPR [6] (*Sparse Unified Part-based Regressor*) predstavlja parametarski model ljudskog tijela, tako da se uključuju ekspresije lica, detaljni pokreti šaka i prstiju, kao i detaljni pokreti stopala i nožnih prstiju. MetaHuman Creator [8] predstavlja rješenje zatvorenog tipa, dostupno za upotrebu putem *web* stranice. Pomoću njega se mogu kreirati visoko detaljni digitalni avatari. Upotreba takvih avatara u naučne svrhe je, međutim, značajno ograničena restriktivnom licencom koja onemogućava upotrebu osim na strogo definisane načine. U narednoj sekciji je opisan MakeHuman softverski alat, koji takođe služi za generisanje 3D modela ljudskog tijela na osnovu postavke parametara, a koji je upotrijebljen u ovom radu jer omogućava kreiranje velikog broja primjeraka sa značajnim varijacijama u obliku, pozi i odjeći, a koji su automatski opremljeni sa animacijskim skeletom koji uključuje i zglobove prstiju.

Neki od navedenih skupova stvarnih 3D skenova sadrže tzv. registracije za parametarske modele, što je najčešće SMPL model, a koje predstavljaju liste vrijednosti parametara koje se mogu iskoristiti da se na osnovu šablona parametarskog modela kreira šablonski 3D model koji se poklapa sa 3D skenom. S obzirom na to da parametarski modeli često uključuju i mogućnost izračunavanja lokacija zglobova šablonskog modela na osnovu lokacija površinskih tačaka, moguće je iskoristiti te lokacije kao anotacije za skelet 3D skenova. Međutim, većina skupova podataka nema registracije za parametarske modele koje uključuju tačke prstiju. Postoje softverske skripte za konverziju šablonskih 3D modela između različitih šablonskih modela [52], ali ne postoji garancija da će se, nakon konverzije, šake šablonskog modela, nastalog konverzijom, adekvatno poklapati sa 3D skenovima. Čak i da skup podataka ima registracije za parametarske modele koji uključuju zglobove prstiju, opet je potrebno da i postoje varijacije u obliku i pozi, kao i da postoji dovoljna količina 3D skenova za obučavanje modela mašinskog učenja.

2.3.1 MakeHuman

MakeHuman je besplatan softver sa grafičkim interfejsom, otvorenog tipa, za kreiranje fotorealističnih 3D modela ljudskog tijela na osnovu konfiguracije parametara. Izgled MakeHuman korisničkog interfejsa dat je na Slici 2.2. Korišten je u ovom radu za generisanje skupa podataka, opisanog u Glavi 4.



Slika 2.2: Grafički korisnički interfejs MakeHuman alata.

U nastavku slijedi spisak podesivih parametara u MakeHuman verziji 1.2.1.

Opšti parametri koji se mogu konfigurirati su: pol (100% žensko - neutralno - 100% muško), dob (1-90 godina), istaknutost mišića (0% - 100%), masa (50% - 150%), visina (~125 cm - ~240 cm), proporcije tijela (0% - 100%) i etnicitet.

Parametri oblika tijela specifični za muški pol su:

- komponenta za ovalni, *apple*, oblik tijela koja povećava zaobljenost trbuha,
- komponenta za pravougaonost tijela čijim povećanjem se teži da prsa, kukovi i struk imaju istu širinu,
- komponenta za usko mršav, *lean column*, oblik čijim povećanjem se takođe teži da prsa, kukovi i struk imaju istu širinu, uz mršaviju građu,
- komponenta za trouglast oblik tijela, čijim povećanjem se sužavaju prsa i povećava težina donjeg dijela tijela,
- komponenta za obrnuto trouglast oblik tijela, čijim povećanjem se naglašavaju mišići, šire ramena i sužavaju kukovi i

- komponenta za trapezoidni, atletske, oblik, čijim povećanjem se postiže da je gornji dio tijela veći od donjeg dijela tijela.

Parametri oblika tijela specifični za ženski pol su:

- komponenta za *apple* oblik tijela čijim povećanjem se smanjuje naglašenost struka, prsa i abdomena, te se dobija okruglasta silueta,
- komponenta za dijamant oblik tijela čijim povećanjem kukovi postaju širi od ramena, te se smanjuju definisanost struka i izduženost prsa,
- komponenta za *full hourglass* oblik tijela čijim povećanjem kukovi postaju približno jednaki prsima po veličini, uz jasno definisanu užu liniju struku i okruglost kukova,
- komponenta za *neat hourglass* mršaviji oblik tijela čijim povećanjem kukovi postaju približno jednaki prsima po veličini, uz jasno definisanu užu liniju struku,
- komponenta za obrnuto trouglast oblik tijela, čijim povećanjem se šire ramena i prsa, sužavaju kukovi, te struk postaje manje izdefinisan,
- komponenta za usko mršav oblik (*lean column*) čijim povećanjem se teži da prsa, kukovi i struk imaju istu širinu,
- komponenta za pravougaonost tijela čijim povećanjem se teži da prsa, kukovi i struk imaju istu širinu i
- komponenta za trouglast oblik tijela, čijim povećanjem se smanjuju širina ramena i obim struka, te se povećava širina kukova.

Parametri oblika tijela pod uticajem hormonskih efekata, u prilagođenim varijantama za muški i za ženski pol su:

- komponenta za produkciju kortizola (hormon stresa) od strane nadbubrežnih žlijezda, koja bi trebala da kontroliše debljinu osobe koja proističe iz pretvaranja, usljed prevelike produkcije kortizola, proteina potrebnih za mišiće u šećer,
- komponenta za efekte oštećenja jetre, koja povećava fluidni izgled trbuha,
- komponenta koja modeluje uticaj na oblik tijela muškog pola usljed visoke inzulinske rezistencije, niskog nivoa testosterona i visokog nivoa estrogena koji se ne može pročistiti iz jetre,
- komponenta koja modeluje uticaj na oblik tijela ženskog pola usljed viška estrogena što dovodi do povećanja debljine dijelove tijela u kojima se proizvodi, kao i okolnim dijelovima tijela, trbuhu, kukovima i butinama, kao i
- komponenta koja modeluje hipofunkciju štitne žlijezde zbog koje se usporava metabolizam i povećava gojaznost.

Konfigurabilni parametri lica su:

- oblik: starost, debljina obraza, ugao brade, ovalnost, okruglost, pravougaonost, sličnost kocki, sličnost trouglu, sličnost obrnutom trouglu, sličnost obliku dijamanta i ispupčenost parijetalne regije,
- veličina: dubina, širina, visina, horizontalni pomjeraj, vertikalni pomjeraj i dubinski pomjeraj (po sagitalnoj osi),
- parametri čela: izbočenost, visina, kranijalni oblik i izbočenost sljepoočnica,
- parametri obrva: izbočenost, ugao, visina
- parametri vrata: istaknutost podbratka, dubina, širina, visina, horizontalni pomjeraj, vertikalni pomjeraj, dubinski pomjeraj (po sagitalnoj osi) i debljina potiljka
- parametri posebni za oba oka: volumen podočnjaka, distorzija podočnjaka, visina podočnjaka, ugao ocnog nabora, prisutnost epikantusa, volumen ocnog nabora, pozicija ocnog nabora, otvorenost oka na dijelu do nosa, otvorenost oka na sredini, otvorenost oka kod vanjskog ugla, horizontalni pomjeraj vanjskog ugla, horizontalni pomjeraj unutrašnjeg ugla, horizontalni pomjeraj čitavog oka, vertikalni pomjeraj čitavog oka, veličina oka, vertikalni pomjeraj vanjskog ugla, vertikalni pomjeraj unutrašnjeg ugla
- parametri za veličinu nosa: vertikalni pomjeraj, dubinski pomjeraj (po sagitalnoj osi), horizontalni pomjeraj, visina, širina i dužina,
- parametri za detalje veličine nosa: širina nosnica, širina vrha nosa, visina na kojoj se nalazi vrh nosa, širina gornjeg dijela nosa, širina središnjeg dijela nosa, širina donjeg dijela nosa
- karakteristike nosa: spljoštenost (kompresija), kvrgavost nosa, udubljenost gornjeg dijela nosa, grbavost nosa, volumen vrha nosa, ugao nosnica, vertikalni pomjeraj vrha nosa, ugao septuma, otvorenost nosnica,
- karakteristike veličine usta: širina usana, vertikalna veličina usana, spljoštenost usana, horizontalni pomjeraj usana, vertikalni pomjeraj usana i dubinski pomjeraj (po sagitalnoj osi) usana,
- karakteristike detalja veličine usta: vertikalna veličina donje usne, širina donje usne, vertikalna veličina gornje usne, širina gornje usne i širina Amorovog luka usne,
- karakteristike usta: naglašenost rupica, naglašenost linija osmijeha, zakrivljenost donje usne, vertikalni pomjeraj uglova usana, veličina središnjeg dijela donje usne, volumen donje usne, volumen usnenog žlijeba (filtruma), volumen gornje usne, zakrivljenost gornje usne, veličina središnjeg dijela gornje usne i zakrivljenost Amorovog luka usne,

- parametri posebni za oba uha: pomjeraj po dubini, veličina, vertikalni pomjeraj, visina, veličina resice, oblik (šiljasto-trokut), rotacija uha, oblik (kockasto-okruglo), širina, ispupčenost čitavog uha ka naprijed (sličnost krilu) i ispučenost gornjeg dijela uha ka naprijed,
- parametri brade i vilice: tonus bočnih dijelova brade, rascjep brade, isticanje brade, širina brade, visina brade, zaobljenost brade i naglašenost prognatizma i
- parametri posebni za oba obraza: vanjski volumen obraza, istaknutost obraza, unutrašnji volumen obraza i vertikalni pomjeraj obraza,

Parametri trupa su: dubina, širina, visina, horizontalni pomjeraj, vertikalni pomjeraj, dubinski pomjeraj (po sagitalnoj osi), sličnost konusnom obliku, istaknutost najšireg leđnog mišića i istaknutost pektoralnog mišića.

Parametri kukova su: dubina, širina, visina, horizontalni pomjeraj, vertikalni pomjeraj, dubinski pomjeraj (po sagitalnoj osi) i vertikalni pozicija struka.

Parametri trbuha su: istaknutost trbušnih mišića (muskularni tonus), sličnost obliku trudnoće, istaknutost pupka i pozicija pupka.

Parametri karličnog regiona su: volumen gluteusa, muskularni tonus pelvisa i volumen perineuma.

Parametri šaka, posebni za obje šake, su: udaljenost između prstiju, prečnik prstiju, dužina prstiju, veličina šake i pozicija šake (izdužuje podlakticu od šake).

Parametri stopala, posebni za oba stopala su: veličina stopala, horizontalni pomjeraj (po frontalnoj osi) i dubinski pomjeraj (po sagitalnoj osi).

Parametri ramena su: mišićna masa lijevog ramena i mišićna masa desnog ramena.

Parametri ruku, posebni za obje ruke su: debljina podlaktice po jednoj osi, dužina podlaktice, debljina podlaktice po drugoj osi, sužavanje podlaktice, mišićna masa podlaktice, debljina nadlaktice po jednoj osi, dužina nadlaktice, debljina nadlaktice po drugoj osi, sužavanje nadlaktice i mišićna masa nadlaktice.

Parametri nogu, posebni za obje noge su: horizontalni pomjeraj koljena, debljina potkoljenice po dubini (po sagitalnoj osi), horizontalna debljina potkoljenice (po frontalnoj osi), sužavanje potkoljenice, mišićna masa potkoljenice, debljina natkoljenice po dubini (po sagitalnoj osi), horizontalna debljina natkoljenice (po frontalnoj osi), vertikalni pomjeraj koljena, sužavanje natkoljenice i mišićna masa natkoljenice.

Parametri koljena su: visina natkoljenica i visina potkoljenica.

Parametri koji se kategorišu kao mjere tijela su: obim vrata, visina vrata, obim nadlaktica, dužina nadlaktica, dužina podlaktica, obim ručnog zgloba, izduženost prsa, obim poprsja, donji

obim prsa, obim struka, dužina od potiljka do struka, dužina od struka do središta kukova, širina ramena, obim kukova, visina natkoljenice, obim butina, obim koljena, visina potkoljenica, obim potkoljenica i obim skočnog zgloba.

Pored navedenih parametara oblika, moguće je i:

- odabrati topologiju poligonske mreže tijela, očiju i zuba 3D modela,
- odabrati i konfigurirati materijal i teksturu za 3D model,
- odabrati odjeću, kape, obuću, frizuru, oblik obrva, oblik trepavica i oblik zuba,
- odabrati animacijski skelet i pozu 3D modela,
- odabrati i konfigurirati facijalnu ekspresiju i
- kreirati sopstvene dodatke po pitanju navedenih stavki ili preuzeti i iskoristiti postojeće, objavljene od strane MakeHuman zajednice.

MakeHuman zajednica omogućava preuzimanje velikog broja dodatnih resursa ove vrste, pod CC0 licencom, kao i objavljivanje sopstvenih, što je značajno za uvođenje varijacija u skupove podataka za modele mašinskog učenja.

Ugrađeni MassProduce modul omogućava kreiranje velikog broja modela sa nasumičnim oblikom tijela, variranjem vrijednosti parametara pomoću Gausove distribucije. Takođe, može se podesiti da se nasumično biraju i odjevni predmeti, materijali i teksture, frizure, izgledi obrva, trepavica i zuba. Kreirani modeli mogu biti opremljeni i specificiranim animacijskim skeletom. Ovaj modul je korišten u ovom radu kako bi se generisao skup podataka za obučavanje i evaluaciju modela mašinskog učenja.

Glava 3

Arhitektura rješenja

Prije razmatranja predložene arhitekture, neophodno je uvesti pojmove konveksne kombinacije i konveksne ljuske. Uvode se sa naredne dvije definicije prema [53], [54].

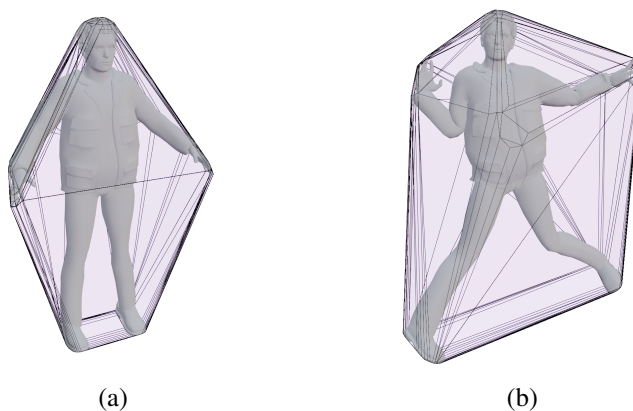
Definicija 8 (Konveksna kombinacija). *Neka je $n \in \mathbb{N}$ i neka su $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$. Ako su $a_1, a_2, \dots, a_n \in \mathbb{R}^{\geq 0}$ i važi $\sum_{i=1}^n a_i = 1$, tada se kaže da je $\sum_{i=1}^n a_i \mathbf{x}_i$ konveksna kombinacija tačaka $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$.*

Definicija 9 (Konveksna ljuska). *Neka je $n \in \mathbb{N}$ i neka su $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$. Skup svih konveksnih kombinacija tačaka $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$,*

$$\left\{ \sum_{i=1}^n a_i \mathbf{x}_i \mid a_1, a_2, \dots, a_n \in \mathbb{R}^{\geq 0} \wedge \sum_{i=1}^n a_i = 1 \right\}, \quad (3.1)$$

naziva se konveksna ljuska tačaka $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$.

Vizuelni primjeri konveksnih ljusaka, relevantni za ovaj rad, dati su na Slici 3.1.



Slika 3.1: Vizualizacija konveksne ljuske na 3D modelima ljudskog tijela: (a) nad modelom u A pozici, (b) nad modelom u nasumično generisanoj pozici.

Kod TARig arhitekture, u formulaciji načina izračunavanja pozicije ključne tačke, (2.21), važi $a_{ij} \in [0, 1]$ i $\sum_{i=1}^{|V|} a_{ij} = 1$. Prema tome, može se zaključiti da je procijenjena lokacija svakog zgloba zapravo konveksna kombinacija ulaznih površinskih tačaka, te da se mora nalaziti unutar njihove konveksne ljuske, što nije iskazano u njihovom radu. Slično važi i za arhitekturu predloženu u ovom radu, kao što će biti objašnjeno u narednoj sekciji.

Predložena arhitektura neuronske mreže

Predložena arhitektura izračunava lokacije ključnih tačaka kao konveksne kombinacije pozicija ulaznih tačaka, slično TARig modulu za šablonske zglobove. Pretpostavlja se, dakle, da ključne tačke treba da budu unutar konveksne ljuske 3D modela. Neuronska mreža procjenjuje koeficijente za te konveksne kombinacije. Varijanta arhitekture sa dinamičkim susjedstvima je data na Slici 3.2. Varijanta arhitekture bez dinamičkih susjedstava je data na Slici 3.3.

Kao osnova, uključena je DGCNN mreža za ekstrakciju obilježja po tački, na osnovu ulaznog skupa tačaka. Ova mreža je izvedena iz implementacije DGCNN *backbone* mreže dostupne u zvaničnom repozitorijumu koda za PointNeXt arhitekturu [31]. Za svaku ulaznu tačku, inicijalna obilježja obuhvataju poziciju tačke i , opcionalno, kao dodatno ulazno obilježje, normalni vektor. Za prvi EdgeConv sloj se izračunavaju susjedstva upotrebom algoritma za određivanje k -najbližih susjeda (kNN) na osnovu Euklidove udaljenosti između tačaka u 3D prostoru. Za naknadne EdgeConv slojeve se izračunavaju grane na osnovu udaljenosti u prostoru obilježja, u varijanti mreže sa dinamičkim susjedstvima, dok se u varijanti mreže bez dinamičkih susjedstava koriste ista susjedstva kao i za prvi EdgeConv sloj.

Kao što je objašnjeno u daljnjem tekstu, DGCNN podmreža je praćena Softmax slojem za primjenu Softmax funkcije za svaki tip ključne tačke. Taj sloj predviđa, za svaku ključnu tačku, koeficijente konveksne kombinacije ulaznih tačaka. Softmax sloj se koristi kako bi se garantovalo da su, za svaku ključnu tačku, koeficijenti konveksne kombinacije nenegativni i da im je suma jednaka 1. Za svaku ključnu tačku, njene koordinate se izračunavaju kao konveksna kombinacija ulaznih tačaka. Izlaz čitave mreže je lista procijenjenih pozicija ključnih tačaka.

Neka važi da su J fiksni broj ključnih tačaka, N broj ulaznih tačaka, $\mathbf{P} \in \mathbb{R}^{N \times 3}$ matrica pozicija ulaznih tačaka, F broj dodatnih ulaznih obilježja, $\mathbf{N} \in \mathbb{R}^{N \times F}$ matrica vektora dodatnih ulaznih obilježja i $\mathbf{F}^{\text{in}} = \begin{bmatrix} \mathbf{P} & | & \mathbf{N} \end{bmatrix} \in \mathbb{R}^{N \times (3+F)}$ matrica ulaznih obilježja ulaznih tačaka, gdje su ulazna obilježja pozicije tačaka i , opcionalno, vektori normala, za koje je $F = 3$. Ako se ne koriste dodatna ulazna obilježja, može se smatrati da je $F = 0$. Najprije se, pomoću EdgeConv slojeva, ekstrahuju obilježja tačaka:

$$\mathbf{F}_1 = \text{EdgeConv}_{N,64}(\mathcal{N}^{\mathbf{P}}, \mathbf{F}^{\text{in}}; \text{MLP}_{N,64}, \text{max}) \in \mathbb{R}^{N \times 64}, \quad (3.2)$$

$$\mathbf{F}_2 = \mathbf{EdgeConv}_{64,64}(\mathcal{N}^{\mathbf{F}_1}, \mathbf{F}_1; \mathbf{MLP}_{64,64}, \max) \in \mathbb{R}^{N \times 64}, \quad (3.3)$$

$$\mathbf{F}_3 = \mathbf{EdgeConv}_{64,128}(\mathcal{N}^{\mathbf{F}_2}, \mathbf{F}_2; \mathbf{MLP}_{64,128}, \max) \in \mathbb{R}^{N \times 128} \text{ i} \quad (3.4)$$

$$\mathbf{F}_4 = \mathbf{EdgeConv}_{128,256}(\mathcal{N}^{\mathbf{F}_3}, \mathbf{F}_3; \mathbf{MLP}_{128,256}, \max) \in \mathbb{R}^{N \times 256}. \quad (3.5)$$

Prethodno navedene jednačine važe kod varijante arhitekture sa dinamičkim susjedstvima. Kod varijante arhitekture bez dinamičkih susjedstava, EdgeConv slojevi iz prethodne tri jednačine koriste isti skup susjeda, $\mathcal{N}^{\mathbf{P}}$, kao i prvi EdgeConv sloj, umjesto skupova $\mathcal{N}^{\mathbf{F}_1}$, $\mathcal{N}^{\mathbf{F}_2}$ i $\mathcal{N}^{\mathbf{F}_3}$.

Obilježja iz EdgeConv slojeva se spajaju i procesiraju dijeljenom podmrežom,

$$\mathbf{F}_5 = \mathbf{MLP}_{2 \cdot 64 + 128 + 256, 512} \left(\left[\mathbf{F}_1 \mid \mathbf{F}_2 \mid \mathbf{F}_3 \mid \mathbf{F}_4 \right]_{N \times (3 \cdot 64 + 128)} \right) \in \mathbb{R}^{N \times 512}. \quad (3.6)$$

Prethodna matrica obilježja se linearnom transformacijom svodi na $N \times J$ matricu,

$$\mathbf{F} = \mathbf{F}_5 \mathbf{L} \in \mathbb{R}^{N \times J}, \quad (3.7)$$

gdje je $\mathbf{L} \in \mathbb{R}^{512 \times J}$ matrica naučenih parametara. Primjenom Softmax operatora po kolonama matrice \mathbf{F} , dobija se matrica pažnje ključnih tačaka prema ulaznim tačkama,

$$\mathbf{A} = \left[\text{Softmax}(\mathbf{F}_{\cdot,1}) \mid \text{Softmax}(\mathbf{F}_{\cdot,2}) \mid \cdots \mid \text{Softmax}(\mathbf{F}_{\cdot,J}) \right] \in \mathbb{R}^{N \times J}. \quad (3.8)$$

Matrica procijenjenih pozicija ključnih tačaka se izračunava kao

$$\hat{\mathbf{J}} = \mathbf{A}^T \mathbf{P} \in \mathbb{R}^{J \times 3}. \quad (3.9)$$

Softmax sloj, dakle, proizvodi matricu $\mathbf{A} \in \mathbb{R}^{N \times J}$ u kojoj k -ta kolona sadrži koeficijente konveksne kombinacije za k -tu ključnu tačku. Pozicija k -te ključne tačke izračunava se, dakle, kao

$$\hat{\mathbf{J}}_{k,\cdot} = \sum_{i=1}^N \mathbf{A}_{i,k} \mathbf{P}_{i,\cdot}. \quad (3.10)$$

Softmax operator garantuje validnost ograničenja

$$(\forall i \in \{1, 2, \dots, N\}) (\forall k \in \{1, 2, \dots, J\}) \mathbf{A}_{i,k} \geq 0 \text{ i} \quad (3.11)$$

$$(\forall k \in \{1, 2, \dots, J\}) \sum_{i=1}^N \mathbf{A}_{i,k} = 1, \quad (3.12)$$

čime se ključne tačke zatvaraju unutar konveksne ljuske ulaznog skupa tačaka.

EdgeConv operator se implementira kao dvodimenzionalni 1×1 konvolucionni operator sa ko-

rakom 1, koji procesira ulazne $N \times K$ mape obilježja, praćen LayerNorm2D [55] slojem za normalizaciju, sa naučenim zasebnim parametrima afine transformacije po kanalu, i LReLU [56] aktivacijom sa negativnim nagibom 0,2.

Dijeljena pod mreža koja procesira konkatenirana obilježja EdgeConv slojeva se implementira kao jednodimenzionalni konvolucioni operator, sa kernelom reda 1 i korakom 1, koji procesira ulazni niz dužine N sa 512 kanala, praćen LayerNorm1D slojem za normalizaciju, sa naučenim zasebnim parametrima afine transformacije po kanalu, i LReLU aktivacijom sa negativnim nagibom 0,2.

U nekim od eksperimenata su, umjesto LayerNorm slojeva, korišteni BatchNorm slojevi. Međutim, ovo se pokazalo kao znatno lošije rješenje za *batch* veličine 1.

Kao funkcija greške odabrana je suma kvadrata Euklidove udaljenosti (SSE) između lokacija stvarnih ključnih tačaka i procijenjenih lokacija ključnih tačaka,

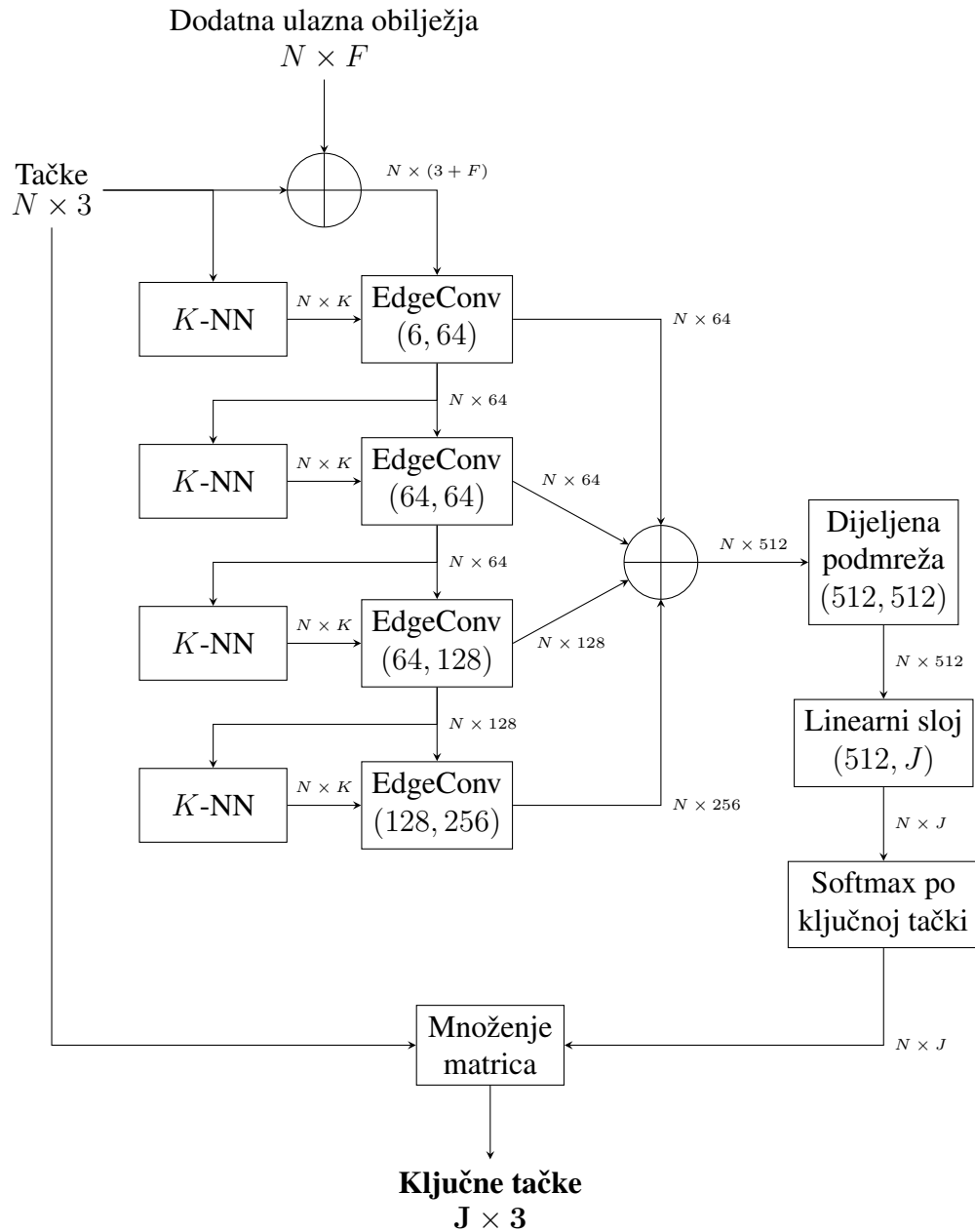
$$L = \sum_{k=1}^J \|\mathbf{j}_k - \hat{\mathbf{J}}_{k,\cdot}\|_2^2, \quad (3.13)$$

gdje je \mathbf{j}_k pozicija k -te stvarne ključne tačke.

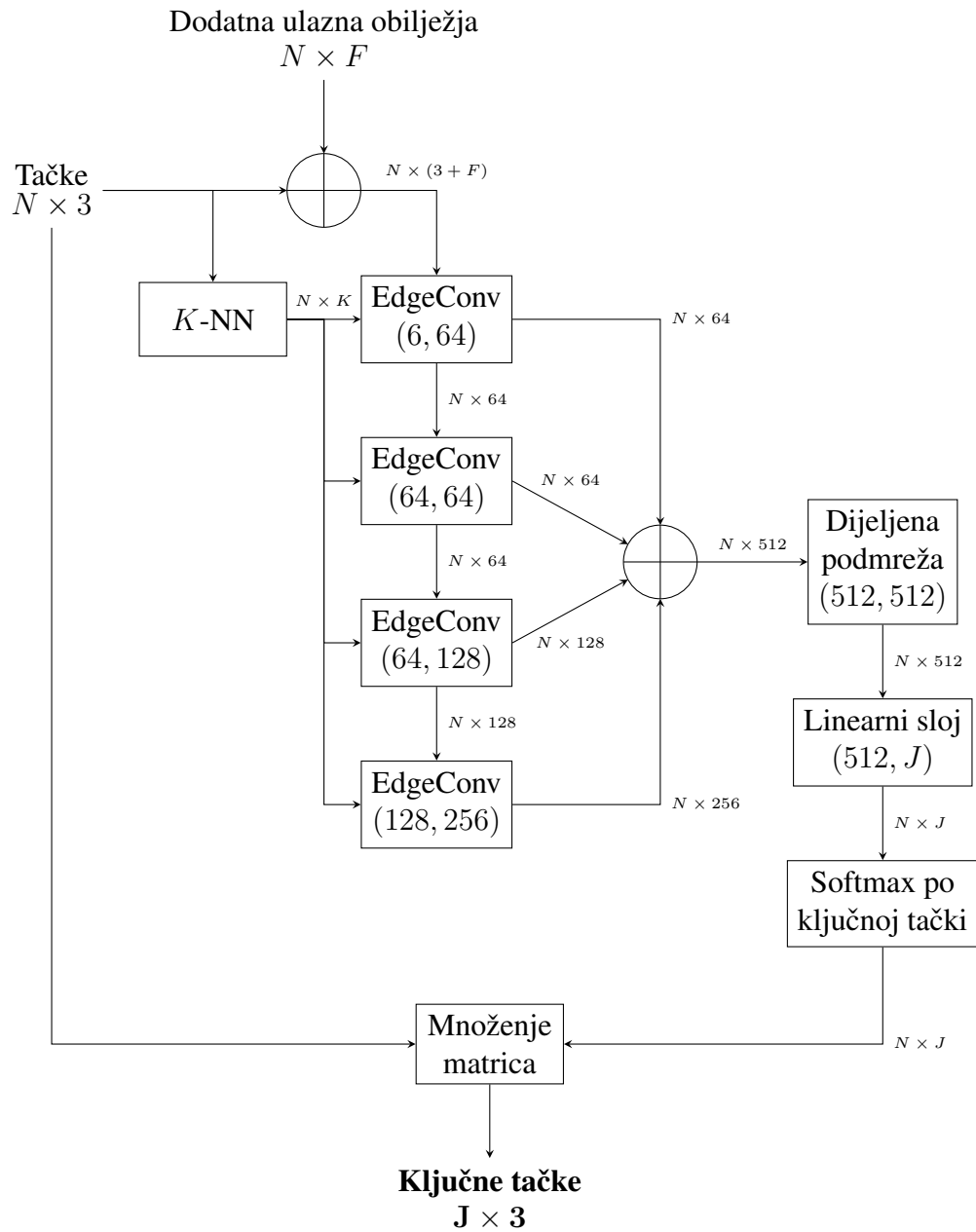
U nekim eksperimentima je korištena, kao funkcija greške, i težinska suma kvadrata Euklidove udaljenosti između lokacije stvarnih ključnih tačaka i procijenjenih lokacija ključnih tačaka,

$$L = \sum_{k=1}^J w_k \|\mathbf{j}_k - \hat{\mathbf{J}}_{k,\cdot}\|_2^2, \quad (3.14)$$

pri čemu je $w_k = 1$ za tačke tijela, odnosno $w_k = 50$ za tačke prstiju.



Slika 3.2: Predložena arhitektura sa dinamičkim susjedstvima. K -NN moduli izračunavaju $N \times K$ matricu indeksa susjeda. Prvi K -NN modul određuje susjede na osnovu udaljenosti u 3D prostoru, dok naknadni K -NN moduli određuju susjede na osnovu udaljenosti u prostoru obilježja ekstrahovanih odgovarajućim EdgeConv modulom.



Slika 3.3: Predložena arhitektura bez dinamičkih susjedstava. K -NN modul izračunava $N \times K$ matricu indeksa susjeda, na osnovu udaljenosti u 3D prostoru.

Glava 4

Skup podataka i evaluacione metrike

4.1 Sintetički podaci

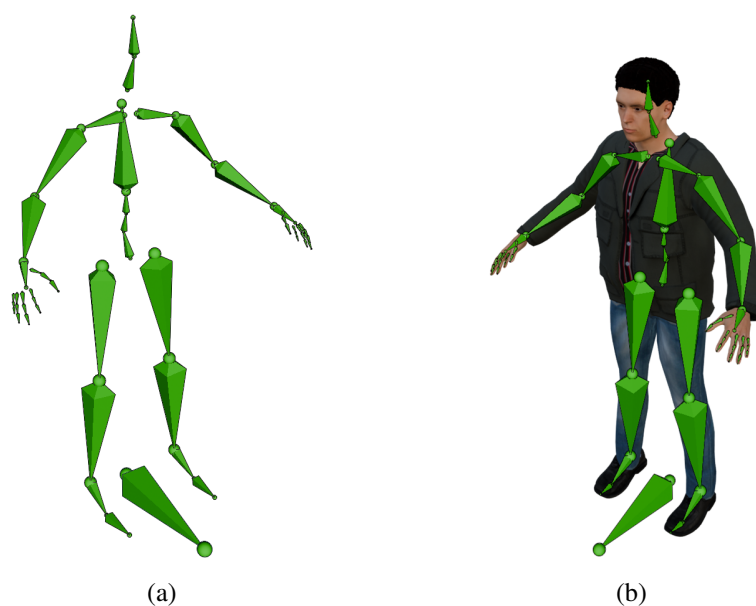
S obzirom na oskudicu dostupnih otvorenih skupova podataka, kao i zatvorenu prirodu komercijalnih skupova podataka, koji se sastoje od kvalitetnih 3D skenova ljudskih tijela sa anotacijama u vidu animacijskih skeleta koji sadrže ključne tačke i tijela i prstiju, u ovom radu se oslanja na pristup u kom se generiše sintetički skup podataka.

Konkretno, koristi se Mass Produce modul MakeHuman softverskog alata [7] za generisanje 3D modela ljudskog tijela. Generisano je 3.000 primjeraka, od čega se 2.400 koristi za obučavanje, 300 za validaciju i 300 za testiranje. Tokom procesa generisanja 3D modela, variraju se parametri oblika, tako da se nasumično određuju korištenjem Gausove raspodjele na način implementiran u Mass Produce modulu. Ovi parametri, koji su detaljno opisani u Glavi 2, uključuju masu, visinu, masu mišića, kao i parametre za individualne dijelove tijela, mijenjajući karakteristike glave, kukova, trupa, ruku, nogu, šaka i prstiju. Takođe, varira se i odjeća na 3D modelima, iako postoje ograničenja po pitanje raznovrsnosti odjeće koja je podrazumijevano dostupna za odabir u MakeHuman okruženju. Primjeri generisanih sintetičkih 3D modela dati su na Slici 4.1. Svi generisani 3D modeli su opremljeni animacijskim skeletom koji se sastoji od 69 ključnih tačaka. Vizualizacija skeleta data je na Slici 4.2. Ovi skeleti predstavljaju stvarne podatke s kojima se predikcije porede u svrhu izračunavanja greške pri obučavanju i evaluaciji.

Iako nisu korišteni ručno formirani animacijski skeleti, upotreba automatski generisanih skeleta je pogodna pri poređenju statističkih modela za automatsko pozicioniranje ključnih tačaka. Ovo proističe iz toga što su takvi skeleti konzistentniji po pitanju pozicioniranja ključnih tačaka u odnosu na površinu 3D modela, kada se uzme u obzir veći broj 3D modela. Ovime se otklanja potreba za uzimanjem u obzir tolerancija zbog ljudske greške.



Slika 4.1: Primjeri generisanih sintetičkih 3D modela sa MakeHuman alatom.



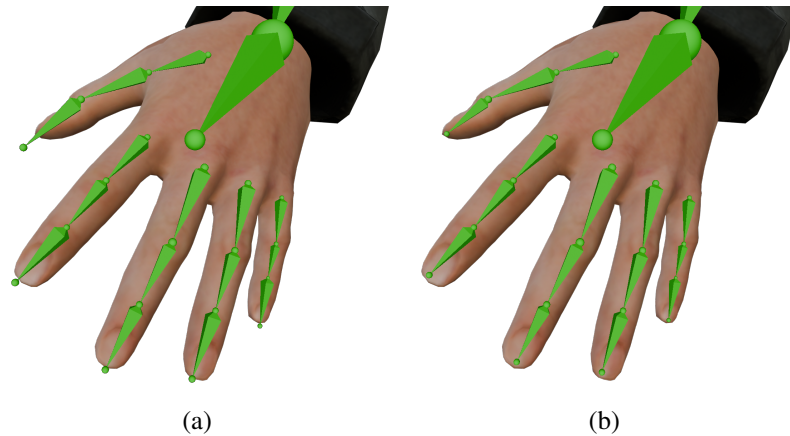
Slika 4.2: Vizualizacija animacijskog skeleta korištenog u generisanom skupu podataka, u početnoj A pozici (a), zajedno sa primjerom 3D modela preko kojega je prikazan skelet (b).

4.2 Korekcija dužina kostiju skeleta

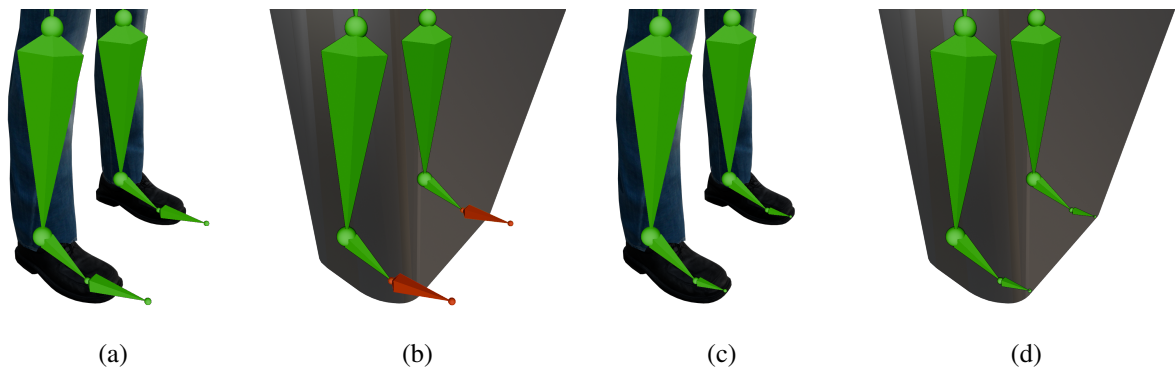
Skeleti iz generisanog skupa podataka sadrže ključne tačke koje se nalaze izvan 3D modela. Ova pojava je većinom zastupljena kod kostiju na dnu skeletalne hijerarhije, kao što su kosti glave, stopala i prstiju, ali takođe i kod kostiju kičme. S obzirom na to da lokalizovane ključne tačke moraju biti unutar konveksne ljuske 3D modela, kao što je objašnjeno u Glavi 3, pogodno je da se stvarne lokacije ključnih tačaka nalaze unutar 3D modela. Ovo je postignuto upotrebom tehnike zasnovanom na emitovanju zraka, pomoću programerskog interfejsa u Blender [57] softverskom paketu.

Za svaku relevantnu kost, emituje se zraka, u smjeru kosti, s ishodištem na poziciji roditeljskog zgloba. Ukoliko je vektor od ishodišta do presjeka zrake i vanjske površine 3D modela kraći od kosti, dužina kosti se smanjuje na 95% dužine tog vektora. Ovime se ne utiče na validnost

animacijskog skeleta, s obzirom na to da su smjer i ishodište kosti nepromijenjeni. Sama dužina kosti se može prilagoditi, prema specifičnim zahtjevima, tokom postprocesiranja. Primjeri su prikazani na Slikama 4.3 i 4.4.



Slika 4.3: Prsti prije (a), i poslije (b), skraćivanja kostiju.



Slika 4.4: Stopala prije skraćivanja kostiju (a), sa prikazom kako kosti onda probijaju konveksnu ljusku 3D modela (b) i stopala poslije skraćivanja kostiju (c), sa prikazom kako su kosti onda unutar konveksne ljuske 3D modela (d).

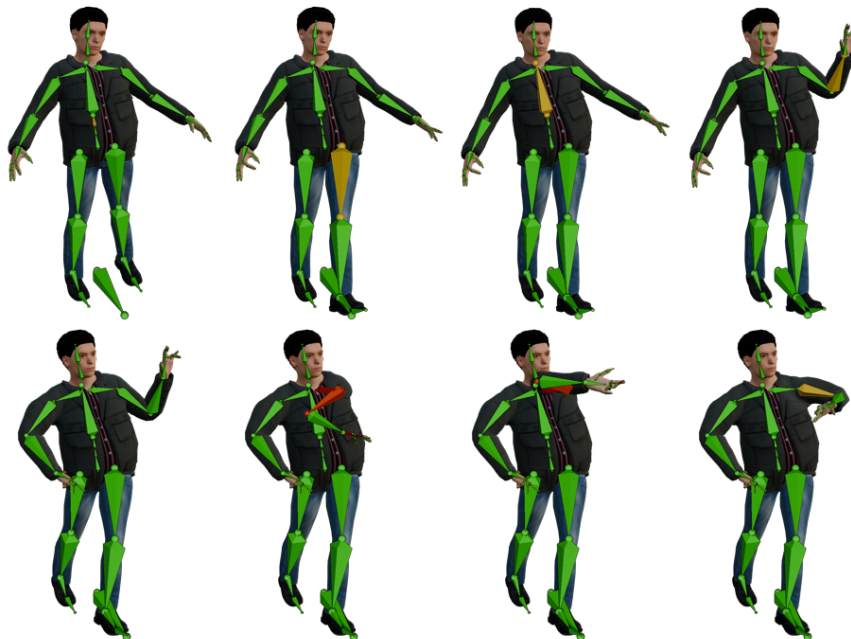
4.3 Generisanje nasumične poze

Generisani 3D modeli su inicijalno pozicionirani u A pozi. Jedan od ciljeva ovog rada jeste evaluacija robusnosti neuronske mreže s obzirom na deformacije usljed varijacije poze. Iz tog razloga se poza 3D modela modifikuje upotrebom animacijskog skeleta, na način da se nasumično rotiraju kosti unutar predefinisanih opsega do 15 stepeni po osi, korištenjem programerskog interfejsa Blender softverskog paketa. Kostii se rotiraju u nasumičnom redosljedu svaki put, tako da se kandidatske orijentacije, prije prihvatanja rotacije, testiraju na kolizije. S obzirom na kompleksnost problema provjere da li je došlo do kolizije, upotrijebljen je efikasan algoritam

zasnovan na emitovanju zraka. Nakon svake rotacije jedne kosti, emituju se zrake od roditeljskih zglobova svih kostiju, u smjeru respektivnih kostiju. Za svaku kost se provjerava da li je došlo do promjene broja prepreka pogođenih takvom zrakom. Ukoliko jeste, kandidatska orijentacija se odbacuje i vrši se iznova nasumična rotacija prvobitne kosti, praćena navedenom provjerom. Primjeri generisanih sintetičkih 3D modela, nakon variranja poze, dati su na Slici 4.5. Na Slikama 4.6 i 4.7 prikazana su stanja 3D modela tokom transformacije poze.



Slika 4.5: Primjeri sintetičkih 3D modela sa Slike 4.1, nakon generisanja nasumične poze.



Slika 4.6: Ilustracija međukoraka transformacije poze rotacijama pojedinačnih kostiju tijela (žuto). U drugom redu je prikazan primjer slučajeva kada se rotacije generišu više puta zbog detektovanih kolizija (crveno).

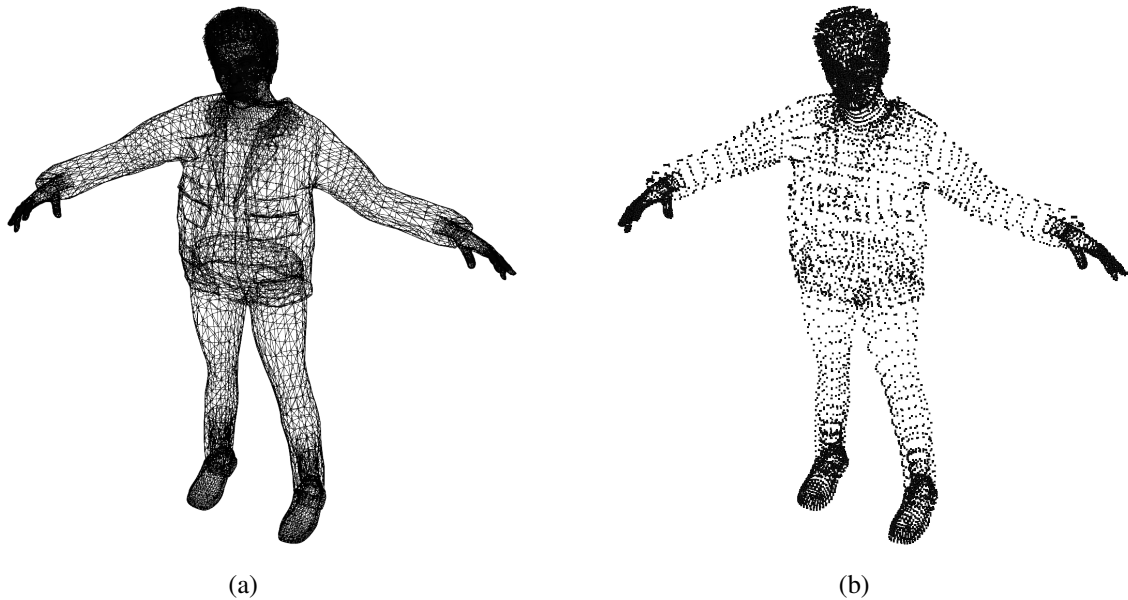


Slika 4.7: Ilustracija međukoraka transformacije poze rotacijama pojedinačnih kostiju prstiju (žuto).

4.4 Rekonstrukcija topologije i normalizacija koordinata

Da bi se smanjili vrijeme izračunavanja i zahtjevi za resursima, 3D model se pojednostavljuje korištenjem MeshLab [58] implementacije algoritma za kolabiranje grana grafa opisanog u [59]. Na taj način su smanjeni broj poligona 3D modela, na 20.480, kao i površinskih tačaka, na opseg između 10.000 i 12.000. Kako bi se postigla invarijantnost na skalu, 3D model se skalira tako da je obuhvaćen jediničnom kockom. Ovo se postiže dijeljenjem pozicija površinskih

tačkaka sa visinom 3D modela. Takođe, model se translira tako da mu je centar ograničavajuće kocke u ishodištu koordinatnog sistema. 3D model se zatim konvertuje u skup tačkaka korištenjem Trimesh programerske biblioteke za procesiranje 3D modela [60]. Na Slici 4.8 se može uporediti 3D model u formi poligonske mreže u odnosu na formu skupa tačkaka.



Slika 4.8: Prikaz 3D modela, (a) u prvobitnoj formi poligonske mreže, prije konverzije u skup tačkaka i (b) nakon pojednostavljivanja, u formi skupa tačkaka.

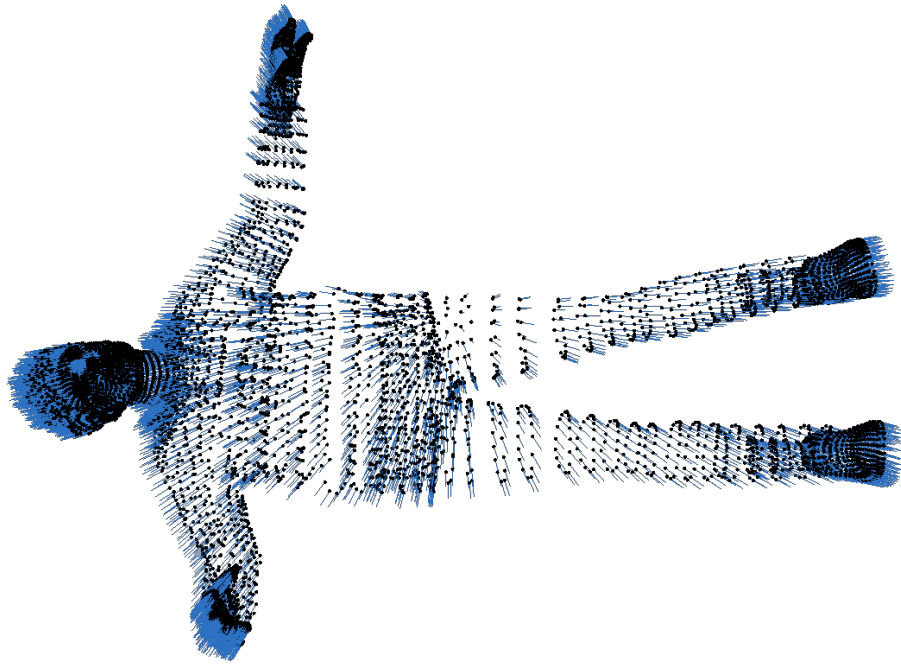
4.5 Estimacija normala

Za estimaciju normala tačkaka u skupu tačkaka, korištena je Open3D biblioteka za procesiranje 3D podataka [61], konfigurisana tako da radi pretragu susjeda hibridnom pretragom koja uzima u obzir do 8 najbližih susjeda sa udaljenošću do 6% visine 3D modela, proizvodeći normale kao na Slici 4.9, koje se koriste kao dodatna ulazna obilježja u nekim od eksperimenata.

4.6 Evaluacione metrike

Korištene su dvije evaluacione metrike: 1) prosječna greška u poziciji po ključnoj tački (*Mean Per Joint Position Error*, MPJPE) i 2) procenat korektnih ključnih tačkaka (*Percentage of Correct Joints*, PCJ).

Greška u poziciji pojedinačne ključne tačke se izračunava kao Euklidova udaljenost između stvarne pozicije te ključne tačke i procijenjene pozicije. MPJPE greška za kategoriju ključnih tačkaka se izračunava kao prosječna vrijednost grešaka u poziciji pojedinačnih ključnih tačkaka koje pripadaju toj kategoriji. MPJPE metrika je slična CD-J2J metrici iz RigNet i TARig radova,



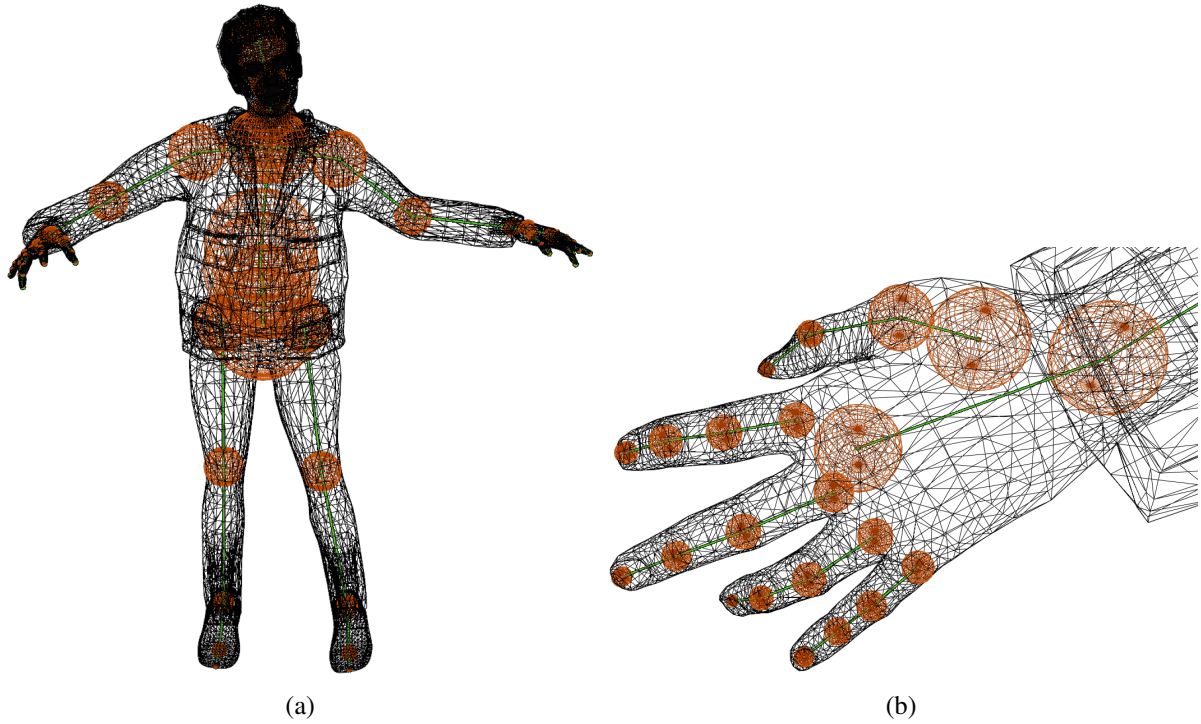
Slika 4.9: Prikaz estimacija vektora normala.

s obzirom na to da se izračunavaju apsolutne udaljenosti između predikcija i stvarnih ključnih tačaka.

Da bi se ključna tačka smatrala kao korektno lokalizovana, prema PCJ metrici, udaljenost između procijenjene pozicije i stvarne pozicije treba biti unutar unaprijed izračunatog praga udaljenosti. Da bi se ovaj prag izračunao za pojedinačnu ključnu tačku, za svaku kost koja sadrži zglob predstavljen tom ključnom tačkom, emituju se zrake okomite na tu kost. Ishodište tih zraka je u stvarnoj poziciji ključne tačke, dok se rotiranjem zrake oko kosti, u opsegu $[0, 2\pi)$, uniformno pokrivaju različiti mogući smjerovi. Ako sve zrake promaše površinu 3D modela, ova akcija se ponavlja, s tim da se ishodište zrake pomjera ka roditeljskom zglobu. Kada se odrede sve tačke presjeka između zraka i 3D modela, izračunava se prosječna udaljenost između stvarne ključne tačke i bliže polovine tačaka presjeka. Ova prosječna udaljenost se može smatrati lokalnim poluprečnikom oblika [33], [62]. Ona se skalira sa faktorom praga udaljenosti u intervalu $[0, 1]$, te se dobijena vrijednost koristi kao prag udaljenosti od stvarne ključne tačke unutar kojeg lokalizovana ključna tačka mora biti da bi se smatrala tačnom. Predložena PCJ metrika je slična IoU metrici iz RigNet i TARig radova, s obzirom na to da se izračunava procenat tačnih procjena, gdje je uslov za tačnost određen pragom udaljenosti izračunatim na osnovu lokalnog prečnika oblika [33], [62].

Na Slici 4.10 data je, na jednom primjeru 3D modela, vizualizacija sfera, čiji je prečnik jednak izračunatom lokalnom prečniku oblika, unutar kojih se odgovarajuće lokalizovane ključne tačke moraju nalaziti, za faktor praga udaljenosti jednak 1, da bi se smatrale tačnim. Vidljivo je da su

tada pragovi za neke ključne tačke restriktivniji od drugih, na primjer za ključne tačke prstiju, što znači da za njih veliki značaj imaju i tačnosti pri uzimanju većih faktora praga udaljenosti. S druge strane, lokalni prečnici oblika za tačke kičme su značajno veći, što znači da je smisljeno zahtijevati da tačnost bude dobra za niže faktore praga udaljenosti, jer veći faktori praga udaljenosti dozvoljavaju značajna odstupanja za te tačke.



Slika 4.10: Vizualizacija sfera unutar kojih se lokalizovane ključne tačke trebaju nalaziti kako bi se smatrale tačnim, za tačke tijela (a) i za tačke prstiju (b).

Glava 5

Eksperimentalni rezultati

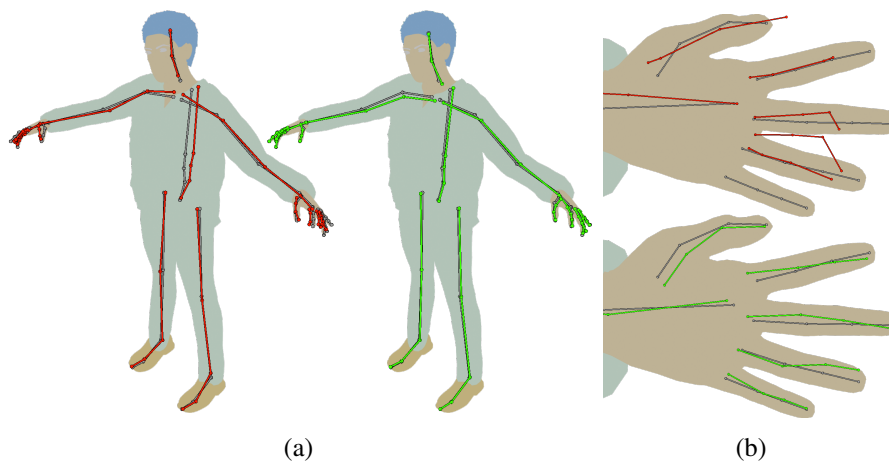
5.1 Pregled eksperimenata

U ovom poglavlju naveden je spisak izvršenih eksperimenata. Mreže obučene prema predloženoj arhitekturi (predložene mreže, u daljnjem tekstu), na skupu podataka opisanom u prethodnoj Glavi, upoređene su i međusobno, a i sa TARig modulom za šablonske zglobove (*template joint module*). S obzirom na to da izvorni kod za TARig [34] nije objavljen, reimplementiran je njihov modul za šablonske zglobove (TARig-TJM), te je i on, takođe, obučen i evaluiran na istom skupu podataka. Kako su uočene sličnosti između arhitektura, TARig-TJM reimplementacija je zasnovana na javno dostupnom RigNet [33] repozitorijumu programskog koda. Izvršeni su sljedeći eksperimenti, za koje su dati nazivi koji će biti upotrebljavani u daljnjem izlaganju:

- TARig-TJM, mreža obučena na osnovu reimplementacije TARig-TJM modula,
- E(BN), predložena mreža sa BatchNorm slojevima,
- E(BN-N), predložena mreža sa BatchNorm slojevima, bez korištenja normalnih vektora,
- E(BN-D), predložena mreža sa BatchNorm slojevima, bez dinamičkih susjedstava,
- E(LN), predložena mreža sa LayerNorm slojevima,
- E(LN-N), predložena mreža sa LayerNorm slojevima, bez korištenja normalnih vektora,
- E(LN-ND), predložena mreža sa LayerNorm slojevima, bez korištenja normalnih vektora i bez dinamičkih susjedstava,
- E(LN+SGP), predložena mreža sa LayerNorm slojevima, u kojoj je kvadratna greška za ključne tačke prstiju skalirana sa 50 i
- E(LN+SGP-ND), predložena mreža sa LayerNorm slojevima, bez korištenja normalnih vektora i bez dinamičkih susjedstava, u kojoj je kvadratna greška za ključne tačke prstiju skalirana sa 50.

Programski kod za obučavanje i evaluaciju mreža zasnovanih na predloženoj arhitekturi je izveden iz repozitorijuma programskog koda koji potiče iz PointMeta [32] i PointNeXt [31] rada, a koji upotrebljavaju PyTorch biblioteku za duboko učenje. Pretprocesiranje podataka, obučavanje TARig–TJM reimplementacije, obučavanje predloženih mreža sa BatchNorm normalizacijom i evaluacija na testnom skupu su izvršeni na računaru sa jednom NVIDIA RTX 4070 12 GB VRAM grafičkom karticom, AMD Ryzen 3600 procesorom i 16 GB RAM memorije. Obučavanje predloženih mreža sa LayerNorm normalizacijom je izvršeno na računaru sa jednom NVIDIA RTX 3090 24GB VRAM grafičkom karticom, Intel i7-9700 procesorom i 64 GB RAM memorije.

U nastavku je pokazano da su svi predloženi metodi bolji u odnosu na TARig–TJM reimplementaciju. Primjer procjena ključnih tačaka sa mrežom iz eksperimenta E(BN) i sa TARig–TJM reimplementacijom, za tijelo i za prste, dat je na Slici 5.1. Pomenuta mreža proizvodi bolje procjene u odnosu na TARig–TJM, pogotovo za tačke prstiju.



Slika 5.1: Vizuelno poređenje estimacije ključnih tačaka tijela (a) i prstiju (b). Crvenom bojom su predstavljene TARig–TJM estimacije, dok su zelenom bojom predstavljene estimacije E(BN) modelom. E(BN) model pokazuje blaga poboljšanja primjetna za tačke kičme i desno koljeno, uz značajna poboljšanja za tačke prstiju. TARig–TJM smješta ključne tačke malog prsta unutar domalog prsta, dok ključne tačke domalog prsta smješta izvan 3D modela.

5.1.1 Obučavanje

TARig–TJM reimplementacija je obučavana sa hiperparametrima opisanim u radu koji uvodi TARig arhitekturu [34]. Za hiperparametre i druge detalje koji nisu opisani u samom radu, reimplementacija se oslanja na podrazumijevane postavke u RigNet repozitorijumu.

Predložene mreže su obučavane sa sljedećom postavkom hiperparametara:

- *batch* veličina 1,

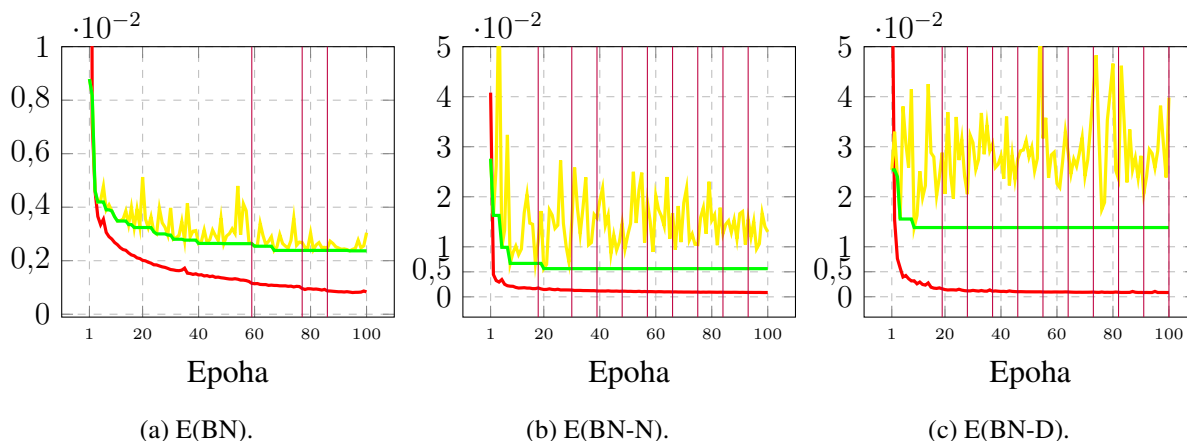
- AdamW [63] optimizator sa opadanjem težine $1e-4$,
- planer koeficijenta učenja s redukcijom na platou validacione greške (ReduceLROnPlateau) sa inicijalnim koeficijentom učenja $1e-3$, strpljenjem od 8 epoha, korakom opadanja od 0,75 i brojem epoha zagrijavanja jednakim 0,
- veličina susjedstava jednaka 80, za EdgeConv module,
- augmentacija skaliranjem tačaka nasumičnom vrijednošću u opsegu od 80% do 120% i
- augmentacija treperenjem (*jitter*) tačaka sa parametrima $\sigma=0,01$ i odsijecanjem na 0,05.

Za predložene mreže sa BatchNorm normalizacijom, obučavanje je trajalo, sa RTX 4070 karticom, oko 8 i po minuta po epohi, odnosno oko 200 ms po ulaznom primjeru. Bez dinamičkih susjedstava, obučavanje je trajalo oko 7 minuta po epohi. Tokovi obučavanja za eksperimente E(BN), E(BN-N) i E(BN-D) predstavljeni su na Slici 5.2. Najbolja prosječna SSE greška nad validacionim skupom je $2,37e-3$ u epohi 94, za E(BN) eksperiment, sa prosječnom SSE greškom na testnom skupu, za epohu 94, jednakom $1,99e-3$. Vidljivo je da je obučavanje nestabilno, s obzirom na oscilacije validacione greške, što se pogoršava ukoliko se ne koriste normale. Još veće pogoršanje vidljivo je kod eksperimenta bez dinamičkih susjedstava, E(BN-D). Prema tome, moglo bi se zaključiti da normale i dinamička susjedstva stabilizuju tok obučavanja i poboljšavaju rezultate. Međutim, kod eksperimenata sa LayerNorm normalizacijom ovo ne važi, što će biti pokazano u nastavku.

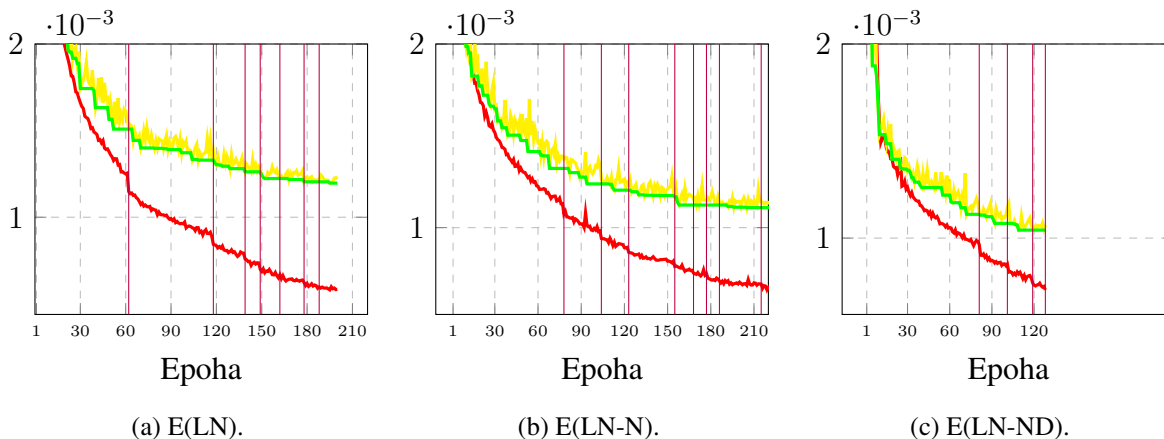
Za predložene mreže sa LayerNorm normalizacijom, obučavanje je trajalo oko 12 minuta i 20 sekundi po epohi, sa RTX 3090 karticom, odnosno 300 ms po primjeru iz trening skupa. Bez dinamičkih susjedstava, obučavanje je trajalo oko 11 minuta i 20 sekundi po epohi. Tokovi obučavanja za eksperimente E(LN), E(LN-N) i E(LN-ND) predstavljeni su na Slici 5.3. Za eksperiment E(LN), najbolja SSE validaciona greška je $1,20e-3$ u epohi 195, a prosječna SSE greška na testnom skupu, za epohu 195, jednaka je $1,06e-3$. Za eksperiment E(LN-N), najbolja SSE validaciona greška je $1,10e-3$ u epohi 205, a prosječna SSE greška na testnom skupu, za epohu 195, jednaka je $1,00e-3$. Prema tome, može se zaključiti da se bez korištenja normala dobijaju bolji rezultati, ako se koristi LayerNorm normalizacija. Za eksperiment E(LN-ND), najbolja SSE validaciona greška je $1,00e-3$ u epohi 109, a prosječna SSE greška na testnom skupu, za epohu 195, jednaka je $0,94e-3$. Pored toga što se bez normala i dinamičkih susjedstava dobijaju bolji rezultati, uočljivo je i da kod eksperimenta bez dinamičkih susjedstava, E(LN-ND), obučavanje mnogo brže konvergira, tako da je potreban znatno manji broj epoha da se dobije manja greška nego kod eksperimenata E(LN) i E(LN-N).

Za predložene mreže sa LayerNorm normalizacijom, u kojima je SSE greška prstiju pomnožena sa 50, dati su, na Slici 5.4, grafici MPJPE greške za tačke prstiju, nad validacionim skupom, u zavisnosti od epohe obučavanja. Vidljivo je da i u ovom slučaju varijanta bez dinamičkih

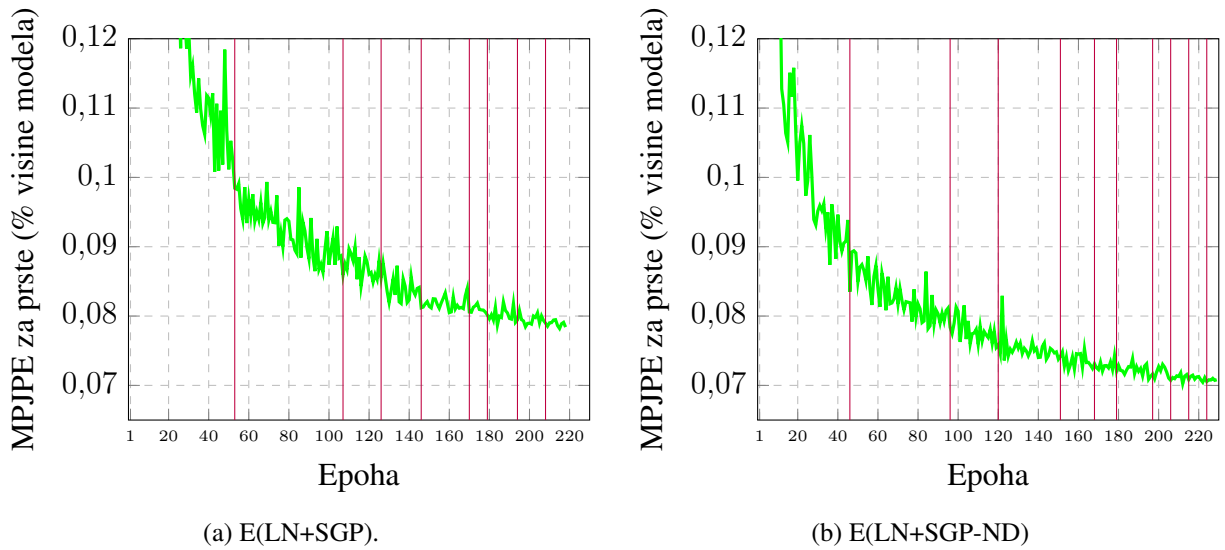
susjedstava, E(LN+SGP-ND), pokazuje bolje rezultate, s tim da obučavanje može da traje znatno duže nego kod E(LN-ND), ako je cilj minimizovati grešku za tačke prstiju. Za eksperiment E(LN+SGP), najbolja neskaliрана SSE validaciona greška je $1,23e-3$ u epohi 210, te je prosječna neskaliрана SSE greška na testnom skupu, za epohu 210, jednaka $1,11e-3$. Za eksperiment E(LN+SGP-ND), najbolja neskaliрана SSE validaciona greška je $1,01e-3$ u epohi 187, te je prosječna SSE greška na testnom skupu, za epohu 187, jednaka $0,95e-3$.



Slika 5.2: Tokovi obučavanja za eksperimente sa BatchNorm normalizacijom: (a) E(BN), (b) E(BN-N) i (c) E(BN-D). Crvena kriva predstavlja SSE grešku tokom obučavanja, žuta kriva predstavlja validacionu SSE grešku, dok zelena kriva predstavlja najbolju postignutu validacionu SSE grešku do trenutne epohe. Ljubičaste vertikalne linije naznačavaju promjene koeficijenta učenja.



Slika 5.3: Tokovi obučavanja za eksperimente: (a) E(LN), (b) E(LN-N) i (c) E(LN-ND). Crvena kriva predstavlja SSE grešku tokom obučavanja, žuta kriva predstavlja validacionu SSE grešku, dok zelena kriva predstavlja najbolju postignutu validacionu SSE grešku do trenutne epohe. Ljubičaste vertikalne linije naznačavaju promjene koeficijenta učenja.



Slika 5.4: MPJPE greška za ključne tačke prstiju na validacionom skupu, u toku obučavanja, (a) Za eksperiment E(LN+SGP) i (b) za eksperiment E(LN+SGP-ND).

5.1.2 Evaluacija

Prosječna greška po ključnoj tački

Rezultati poređenja metoda korištenjem MPJPE metrike su prikazani u Tabeli 5.1, sumarno, a i za pojedinačne kategorije tačaka tijela, kao u i Tabeli 5.2, za pojedinačne prste. Greške su mjerene nakon normalizacije, te predstavljaju postotke visine 3D modela. Za svaku kategoriju ključne tačke, predložene mreže proizvode manje vrijednosti greške u poređenju sa TARig–TJM reimplementacijom. Vidljivo je da, među izvršenim eksperimentima, E(BN-N) i E(BN-D) imaju najveće greške, a zatim E(BN). Korištenjem LayerNorm normalizacije, greške se značajno smanjuju. E(LN-ND) ima najmanju grešku za tačke tijela, s tim da vrijedi napomenuti da u nekim kategorijama tačaka tijela, E(LN-N) i E(LN+SGP-ND) imaju marginalno manju grešku. E(LN+SGP-ND) ima najmanju grešku za sve tačke prstiju.

Tabela 5.1: Poređenja metoda prema MPJPE metrici, na testnom skupu (% visine 3D modela).

Metod	Tijelo	Prsti	Glava	Vrat	Rame	Kičma	Kuk	Lakat	Šaka	Kolj.	Stop.
TARig–TJM	1,47	0,93	1,92	1,76	1,61	1,75	1,63	1,66	0,92	1,87	0,97
E(BN)	0,56	0,19	0,84	0,54	0,53	0,93	0,59	0,45	0,22	0,62	0,43
E(BN-N)	0,77	0,31	1,12	0,81	0,72	1,07	0,78	0,66	0,47	0,90	0,62
E(BN-D)	0,75	0,35	1,03	0,78	0,73	1,12	0,81	0,69	0,35	0,75	0,62
E(LN)	0,40	0,13	0,69	0,32	0,37	0,79	0,44	0,30	0,15	0,42	0,25
E(LN-N)	0,384	0,12	0,66	0,295	0,35	0,79	0,433	0,27	0,142	0,39	0,22
E(LN-ND)	0,375	0,11	0,61	0,297	0,32	0,78	0,435	0,29	0,137	0,36	0,24
E(LN+SGP)	0,43	0,08	0,78	0,37	0,42	0,80	0,48	0,35	0,14	0,47	0,28
E(LN+SGP-ND)	0,385	0,07	0,63	0,287	0,34	0,77	0,44	0,30	0,128	0,40	0,25

Tabela 5.2: Poređenja metoda, za različite kategorije tačaka prstiju, prema MPJPE metrici, na testnom skupu (% visine 3D modela).

Metoda	Palac	Kažiprst	Srednji	Domali	Mali
TARig–TJM	1,04	1,13	0,77	0,84	0,86
E(BN)	0,19	0,17	0,19	0,20	0,19
E(BN-N)	0,28	0,44	0,29	0,28	0,24
E(BN-D)	0,25	0,41	0,41	0,39	0,31
E(LN)	0,12	0,12	0,13	0,13	0,12
E(LN-N)	0,11	0,12	0,125	0,122	0,110
E(LN-ND)	0,10	0,11	0,119	0,123	0,108
E(LN+SGP)	0,071	0,075	0,079	0,083	0,073
E(LN+SGP-ND)	0,065	0,072	0,077	0,079	0,068

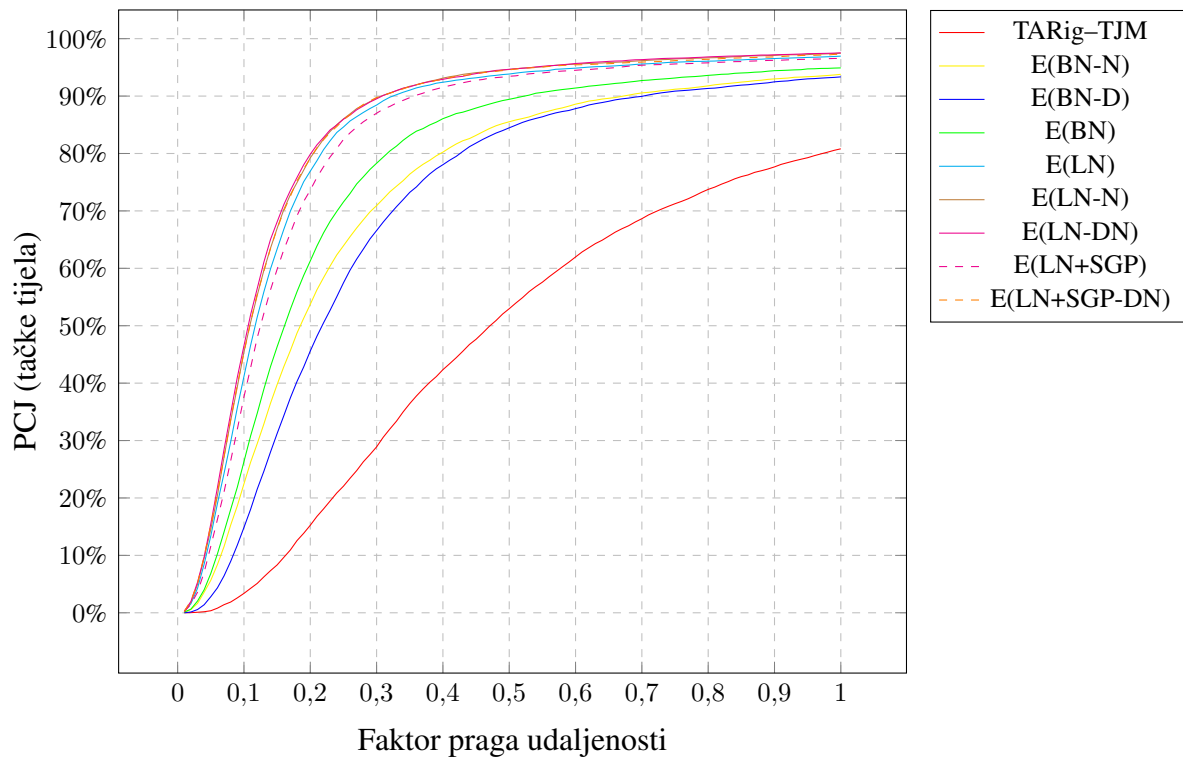
Procenat tačnih ključnih tačaka

Na Slikama 5.5 i 5.6 prikazana su PCJ mjerenja kao funkcija faktora praga udaljenosti, za tačke tijela i za tačke prstiju, respektivno. Vidljiva su značajna poboljšanja svih predloženih mreža u odnosu na TARig–TJM. Vidljivo je da E(BN-D) i E(BN-N) imaju najmanju tačnost, te da nakon njih, po tačnosti, slijedi E(BN). Za tačke tijela, predloženi modeli sa LayerNorm normalizacijom se, po tačnosti, marginalno razlikuju, s tim da je među njima najgori E(LN+SGP), a zatim E(LN). E(LN-N), E(LN-DN) i E(LN+SGP-DN) imaju skoro neznatnu različitost po PCJ metrici za tačke tijela. Za tačke prstiju, E(LN) ima malo manju tačnost u odnosu na E(LN-N) i E(LN-DN), koji se marginalno razlikuju. E(LN+SGP) i E(LN+SGP-DN) imaju znatno bolju tačnost za tačke prstiju od svih ostalih modela, s tim da E(LN+SGP-DN) blago prednjači.

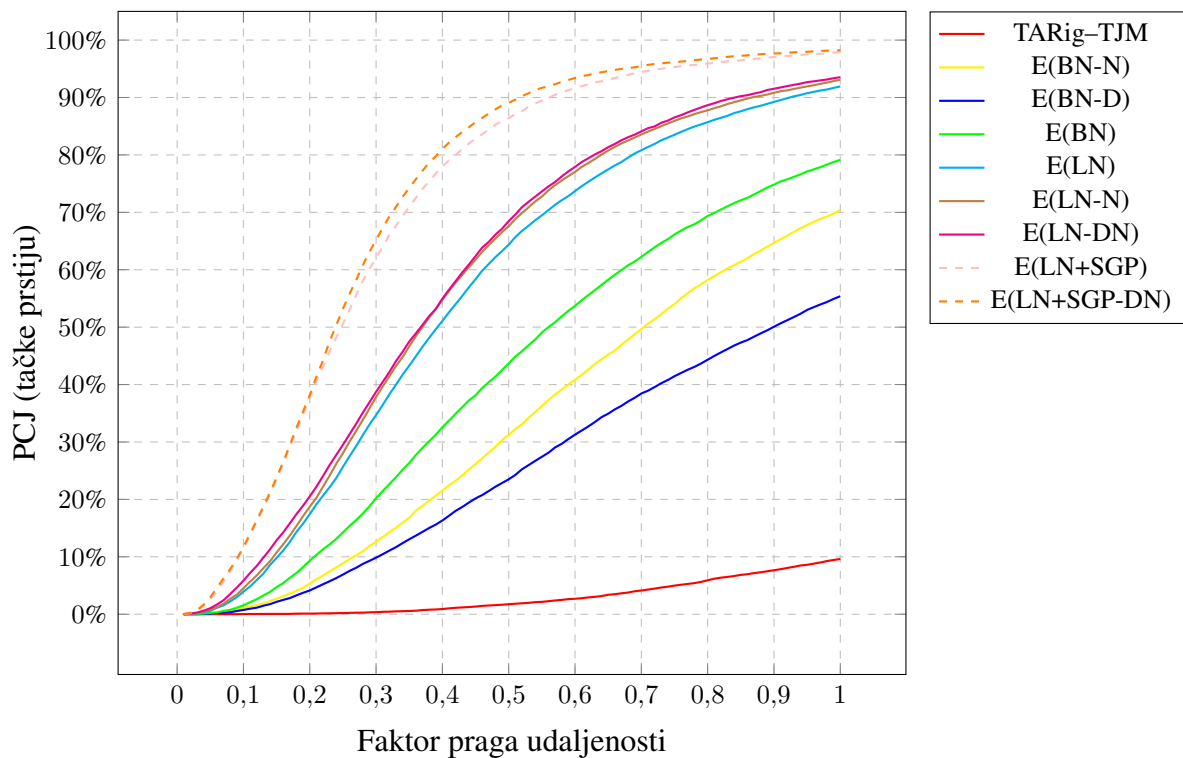
S obzirom na prethodne zaključke, mreža iz eksperimenta E(LN-DN) će se smatrati osnovnim predloženim rješenjem, dok će se mreža iz eksperimenta E(LN+SGP-DN) smatrati predloženim rješenjem za tačke prstiju. Ipak, treba napomenuti da su ove dvije mreže uporedive i za tačke tijela, s tim da se E(LN-DN) može značajno brže obučiti za tu svrhu.

Vrijeme izvršavanja, broj parametara i zauzeće memorije

Predloženi metodi procesiraju 3D model u 2 s. TARig–TJM reimplementacija zahtijeva 25 s, s obzirom na neophodno izračunavanje geodetskih susjeda i mapa pažnje, što zahtijeva 18 s i 6 s, respektivno. Predložena neuronska mreža ima oko 0,4M parametara i zauzima 4,5 MB memorije. TARig–TJM mreža ima 1,8M parametara i zauzima 21 MB memorije, s tim da je većina parametara dijeljena sa drugim TARig modulima, kao što su modul za tok kostiju (*boneflow*) i modul za vezivanje skeleta sa 3D modelom (*skinning*).



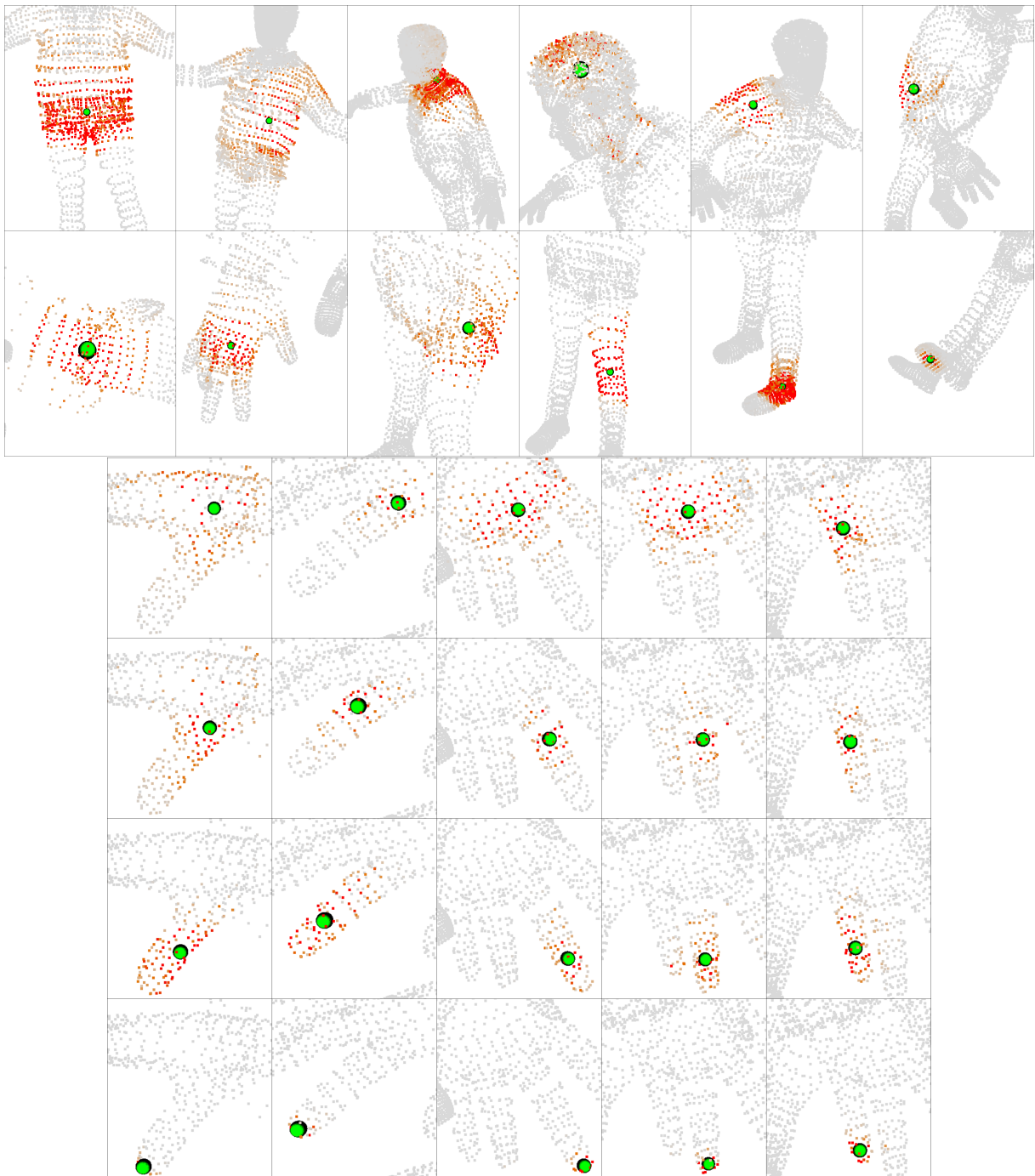
Slika 5.5: Rezultati evaluacije prema PCJ metrici, za tačke tijela na testnom skupu. Predloženi metod pokazuje znatna poboljšanja u lokalizaciji ključnih tačaka tijela.



Slika 5.6: Rezultati evaluacije prema PCJ metrici za tačke prstiju, na testnom skupu. Predloženi metod pokazuje znatna poboljšanje u lokalizaciji ključnih tačaka prstiju.

5.2 Objašnjivost

Na Slici 5.7 je vizuelizovana pažnja za ključne tačke po ulaznim tačkama. Uočljivo je da mreža uči da, za svaku ključnu tačku, posvećuje pažnju površinskim tačkama u njenoj neposrednoj okolini, te da broj površinskih tačaka koje su bitne za određeni tip ključne tačke zavisi od samog tipa ključne tačke.



Slika 5.7: Vizualizacija pažnje ključnih tačaka po ulaznim tačkama.

5.3 Analiza grešaka

U Tabeli 5.3 dati su statistički rezultati za MPJPE grešku za tačke tijela. Značajna disperzija postoji za tačke vrha glave (*head_top*) i tačke vrha stopala (*foot_tip*). Takođe, za te tačke se ističu i maksimalne greške. Kasnijom analizom, u narednoj sekciji, utvrđeno je da ove greške proističu iz neusklađenosti algoritma za skraćivanje kostiju sa skupom podataka.

Tabela 5.3: Statističke informacije o apsolutnoj grešci za ključne tačke tijela, na testnom skupu (% visine 3D modela).

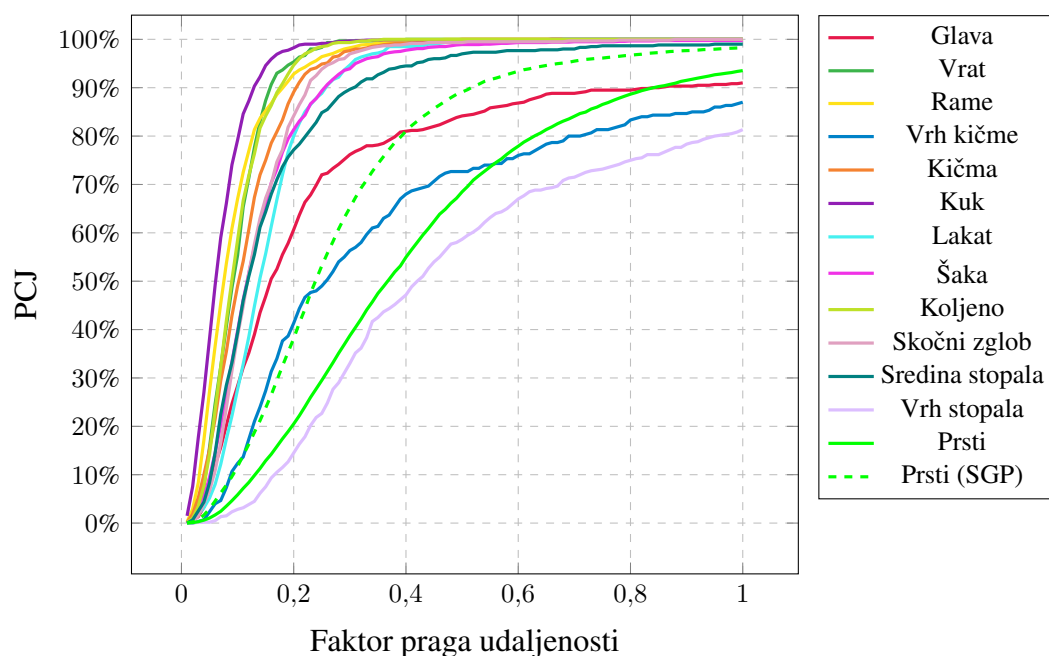
Zglob	Prosjek	Medijan	St. dev.	1. kvartil	3. kvartil	Minimum	Maksimum
pelvis	0,32	0,28	0,19	0,20	0,39	0,03	1,31
spine1	0,51	0,45	0,27	0,32	0,63	0,08	1,93
spine2	0,46	0,39	0,27	0,28	0,58	0,06	2,11
spine3	0,85	0,82	0,38	0,62	1,04	0,13	2,57
spine4	0,94	0,86	0,44	0,63	1,18	0,08	3,10
spine5	0,86	0,74	0,50	0,51	1,11	0,10	3,81
neck	0,30	0,28	0,14	0,19	0,38	0,03	0,86
head	0,27	0,25	0,15	0,15	0,36	0,04	0,89
head_top	0,96	0,56	1,22	0,37	0,93	0,05	7,77
clavicle	0,28	0,26	0,15	0,18	0,36	0,03	1,02
shoulder	0,35	0,33	0,17	0,22	0,46	0,03	1,09
elbow	0,29	0,27	0,17	0,18	0,36	0,03	1,28
wrist	0,12	0,11	0,07	0,08	0,16	0,01	0,47
hand	0,15	0,14	0,08	0,09	0,19	0,02	0,55
thigh	0,46	0,43	0,24	0,28	0,58	0,07	1,88
knee	0,36	0,31	0,21	0,22	0,44	0,02	1,75
ankle	0,22	0,20	0,14	0,14	0,28	0,01	1,63
foot_ball	0,19	0,17	0,13	0,11	0,23	0,02	1,35
foot_tip	0,28	0,21	0,37	0,14	0,30	0,02	3,91

U Tabeli 5.4 dati su statistički rezultati za MPJPE grešku tačaka prstiju. Razmatra se osnovna mreža (Osn.) iz eksperimenta E(LN-DN), kao i varijanta mreže u kojoj je skalirana greška za ključne tačke prstiju sa koeficijentom 50 (SGP), tj. mreža iz eksperimenta E(LN+SGP-DN). Vidljivo je da je, prema većini statistika, E(LN+SGP-DN) varijanta značajno bolja. Maksimalna greška krajnje tačke kažiprsta (*index4*) nije puno bolja, dok su maksimalne greške krajnje tačke domalog prsta (*ring4*) i predzadnje tačke malog prsta (*ring3*) marginalno lošije. Posebno se ističe to da su maksimalne greške za krajnje ključne tačke srednjeg prsta (*mid4*) i malog prsta (*pinky4*) značajno lošije u odnosu na osnovnu varijantu mreže.

Na Slici 5.8 prikazani su rezultati evaluacije predložene neuronske mreže, sa PCJ metrikom za različite tipove ključnih tačaka. Predloženo rješenje proizvodi zadovoljavajuće rezultate za skoro sve ključne tačke. Netačnost za ključne tačke vrha glave i vrha stopala proističe, opet, iz neusklađenosti algoritma za skraćivanje kostiju sa skupom podataka, o čemu će biti riječi u narednoj sekciji. Netačnost za vrh kičme proističe iz prevelike restriktivnosti PCJ metrike, jer se ta tačka često nalazi blizu površine 3D modela, te prag udaljenosti za nju biva malen.

Tabela 5.4: Statističke informacije o apsolutnoj grešci za ključne tačke prstiju, na testnom skupu (% visine 3D modela). Poredi se osnovna varijanta arhitekture (Osn.) sa varijantom u kojoj je skalirana greška za ključne tačke prstiju (SGP).

Zglob	Prosjek		Medijan		St. dev.		1. kvartil		3. kvartil		Minimum		Maksimum	
	Osn.	+SGP	Osn.	+SGP	Osn.	+SGP	Osn.	+SGP	Osn.	+SGP	Osn.	+SGP	Osn.	+SGP
thumb1	0,11	0,07	0,10	0,07	0,07	0,04	0,07	0,05	0,14	0,09	0,00	0,00	0,51	0,37
thumb2	0,10	0,07	0,09	0,06	0,06	0,03	0,06	0,04	0,13	0,09	0,01	0,00	0,52	0,22
thumb3	0,10	0,07	0,09	0,06	0,05	0,03	0,06	0,04	0,13	0,08	0,01	0,01	0,45	0,26
thumb4	0,08	0,05	0,07	0,05	0,04	0,02	0,06	0,03	0,10	0,06	0,02	0,01	0,35	0,20
index1	0,10	0,07	0,09	0,07	0,06	0,04	0,06	0,05	0,13	0,09	0,01	0,00	0,43	0,23
index2	0,11	0,08	0,10	0,07	0,05	0,04	0,07	0,05	0,13	0,10	0,02	0,01	0,34	0,24
index3	0,13	0,08	0,11	0,07	0,07	0,04	0,08	0,05	0,16	0,10	0,01	0,01	0,48	0,27
index4	0,10	0,06	0,08	0,05	0,06	0,04	0,06	0,04	0,12	0,08	0,01	0,00	0,54	0,50
mid1	0,12	0,10	0,11	0,09	0,06	0,04	0,07	0,06	0,15	0,12	0,01	0,01	0,40	0,30
mid2	0,12	0,08	0,10	0,07	0,06	0,03	0,07	0,05	0,15	0,10	0,01	0,01	0,48	0,21
mid3	0,13	0,08	0,12	0,07	0,07	0,04	0,08	0,05	0,17	0,10	0,02	0,01	0,46	0,35
mid4	0,11	0,06	0,09	0,05	0,07	0,06	0,07	0,04	0,13	0,07	0,01	0,01	0,91	1,42
ring1	0,11	0,08	0,10	0,08	0,06	0,04	0,07	0,06	0,14	0,11	0,01	0,01	0,47	0,32
ring2	0,11	0,07	0,10	0,07	0,06	0,03	0,07	0,05	0,14	0,09	0,01	0,01	0,49	0,22
ring3	0,14	0,08	0,12	0,08	0,08	0,04	0,09	0,06	0,17	0,11	0,01	0,01	0,85	0,34
ring4	0,13	0,07	0,12	0,07	0,08	0,05	0,08	0,05	0,16	0,09	0,02	0,01	0,69	0,74
pinky1	0,11	0,08	0,09	0,07	0,07	0,04	0,06	0,05	0,13	0,10	0,01	0,01	0,75	0,35
pinky2	0,11	0,07	0,10	0,06	0,06	0,04	0,07	0,04	0,14	0,08	0,01	0,01	0,56	0,47
pinky3	0,12	0,07	0,10	0,06	0,08	0,04	0,07	0,04	0,14	0,08	0,01	0,01	0,62	0,65
pinky4	0,10	0,06	0,09	0,05	0,07	0,06	0,06	0,03	0,12	0,07	0,01	0,00	0,61	1,21

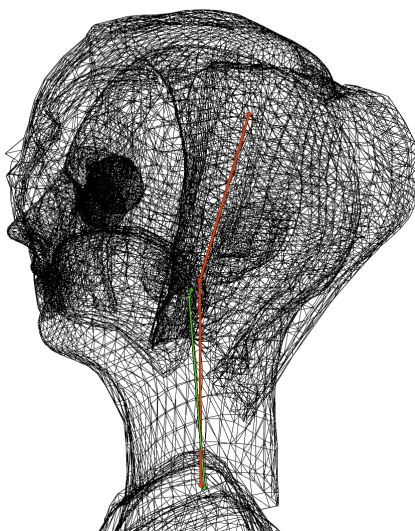


Slika 5.8: Rezultati evaluacije predložene neuronske mreže, sa PCJ metrikom, za različite tipove tačaka tijela. Pod tipom „Šaka” se ne računaju prsti. Pod tipom „Kičma” se ne računa tačka vrha kičme, koja je izdvojena kao poseban tip.

5.4 Vizualna analiza najgorih slučajeva za tačke tijela

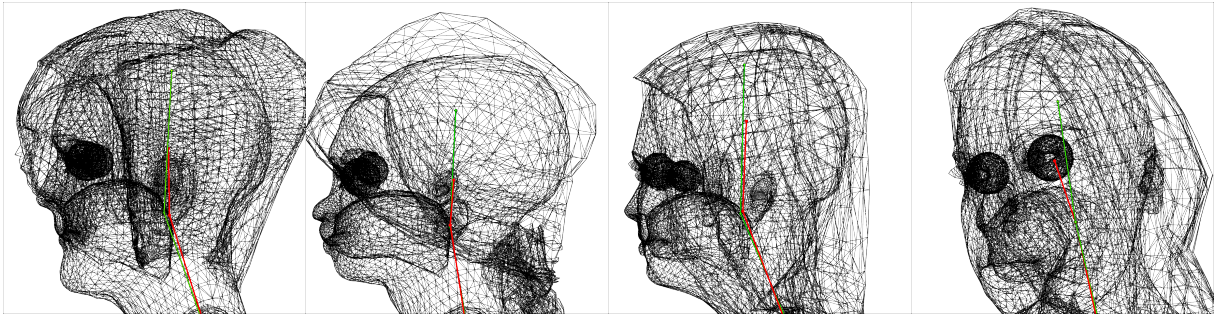
Razmatranjem najvećih grešaka na testnom skupu podataka, filtrirani su modeli sa najvećim greškama. U ovoj sekciji su prikazani primjeri sa najvećim greškama pri lokalizaciji ključnih tačaka. Na svim primjerima, zelenom bojom su predstavljene anotirane kosti, dok su crvenom bojom predstavljene kosti formirane na osnovu procijenjenih lokacija ključnih tačaka.

Na Slici 5.9 prikazan je slučaj sa najvećom greškom za tačke glave. On ukazuje na nedostatak algoritma za skraćivanje kostiju koje izlaze izvan 3D modela, s obzirom na strukturu 3D modela generisanih MakeHuman alatom. Pri skraćivanju kosti glave, zraka emitovana od početka kosti ka njenom kraju, koji bi trebalo da bude pri vrhu glave, pogađa površinu usne duplje, koja je zasebna poligonska mreža u MakeHuman modelima, umjesto vrha glave. S obzirom na to, ta kost se drastično skraćuje, tako da su i njen početak i kraj unutar usne duplje. Ovaj problem je prisutan u 7,6% skupa podataka (23 modela u testnom skupu), te su izmjerene najveće greške upravo na ovakvim slučajevima. Ovaj problem ne bi bio prisutan kod 3D modela koji nastaju 3D skeniranjem, s obzirom na to da oni ne sadrže zasebnu poligonsku mrežu za usnu duplju. Vidljivo je da su procijenjene lokacije bolje u odnosu na pretprocesirane stvarne, što donekle ukazuje na robusnost neuronske mreže s obzirom na nepravilno anotirane podatke.



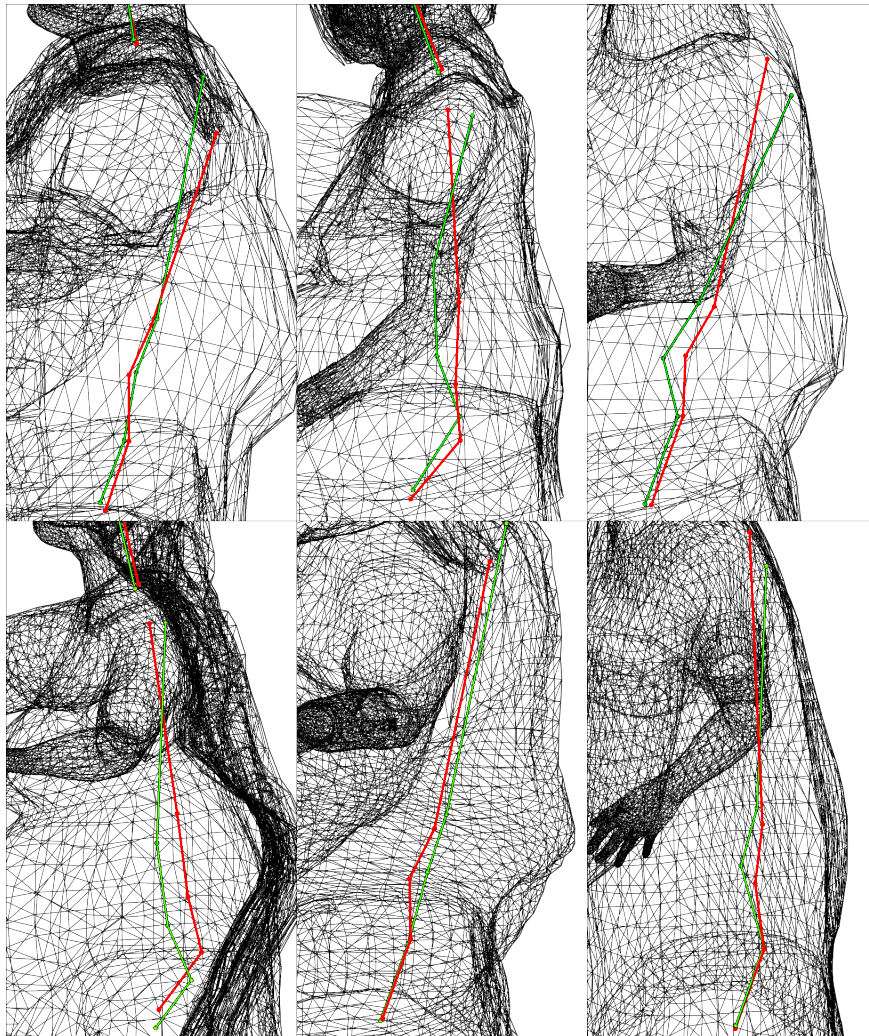
Slika 5.9: Vizualizacija jednog najgorog slučaja za tačke glave koji proističe iz propusta u algoritmu za pretprocesiranje podataka.

Na Slici 5.10 prikazani su i još neki slučajevi sa najvećim greškama za ključne tačke glave. Vidljivo je da je na prvi tri primjera procijenjena kost glave prekratka. U zadnjem primjeru, dužina kosti glave je prekratka i primjetna je greška u uglu, poredeći je sa stvarnom kosti. Pojava ovakvih grešaka može biti uzrokovana, barem djelimično, zbog uticaja nepravilnih anotiranih podataka koji proističu iz greške u pretprocesiranju opisane u prethodnom paragrafu.



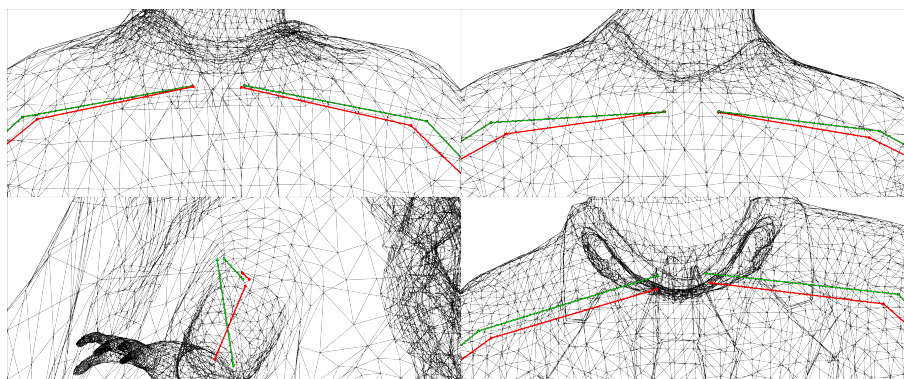
Slika 5.10: Vizualizacija nekih najgorih slučajeva za greške ključnih tačaka glave.

Na Slici 5.11 prikazani su slučajevi sa najvećim greškama za ključne tačke kičme. Vidljivo je da greške za vrh kičme nisu toliko znatne, iako je prema PCJ metrici tačnost za ovaj tip ključne tačke niža u odnosu na ostale tipove tačaka. To može značiti da je PCJ prag za ovaj tip ključne tačke previše restriktivan. Ove greške se mogu smatrati prihvatljivima.



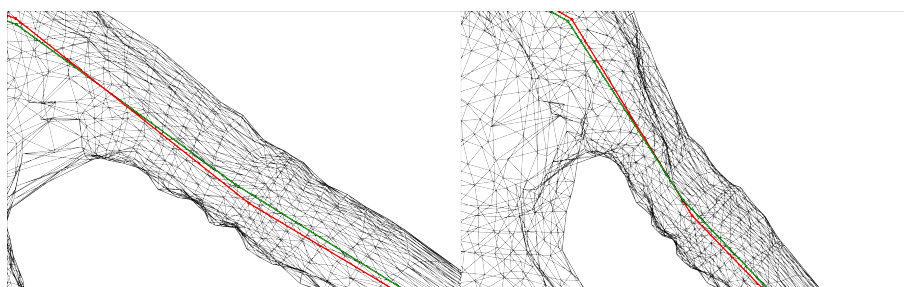
Slika 5.11: Vizualizacija nekih najgorih slučajeva za greške ključnih tačaka kičme.

Na Slici 5.12 prikazani su slučajevi sa najvećim greškama za ključne tačke ramena. Može se vidjeti da se estimacije mogu smatrati tačnim, jer su greške unutar prihvatljivih opsega.



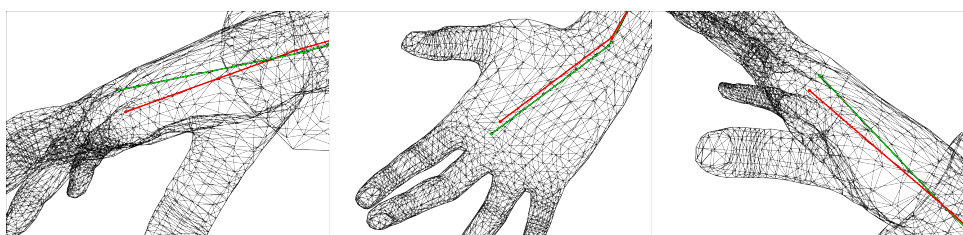
Slika 5.12: Vizualizacija nekih najgorih slučajeva za greške ključnih tačaka ramena.

Na Slici 5.13 prikazani su slučajevi sa najvećim greškama za ključne tačke lakta. Uočljivo je da su odstupanja prihvatljiva na oba primjera.



Slika 5.13: Vizualizacija dva najgora slučaja za greške ključnih tačaka lakta.

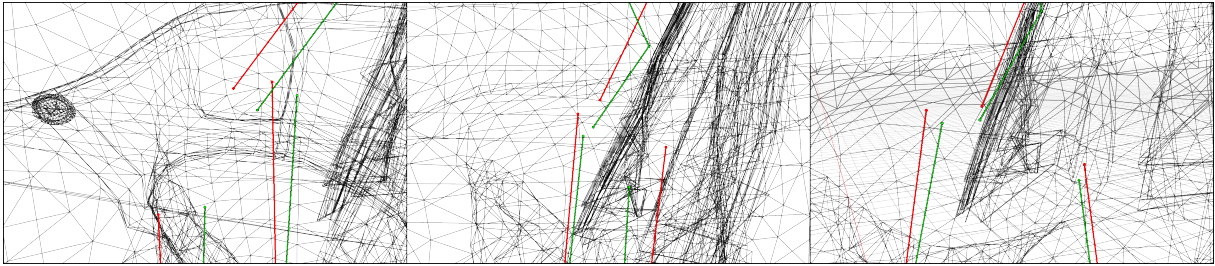
Na Slici 5.14 prikazani su slučajevi sa najvećim greškama za ključne tačke ručnog zgloba i sredine šake. Uočljivo je da su lokacije ovih ključnih tačaka adekvatno procijenjene.



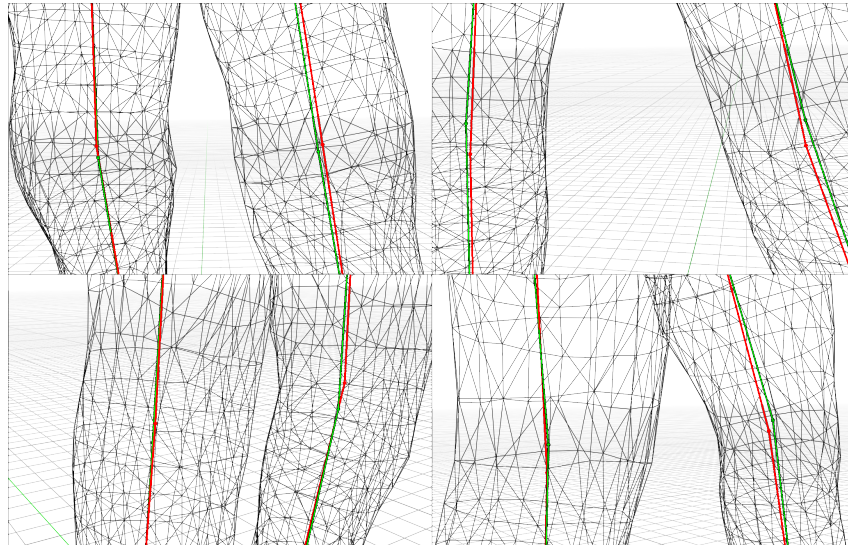
Slika 5.14: Vizualizacija najgorih slučajeva za greške ključnih tačaka šaka.

Na Slici 5.15 prikazani su slučajevi sa najvećim greškama za ključne tačke kukova. Vidljivo je da su odstupanja u prihvatljivom opsegu u svim slučajevima.

Na Slici 5.16 prikazani su slučajevi sa najvećim greškama za ključne tačke koljena. Odstupanja na ovim primjerima su u prihvatljivim opsezima.



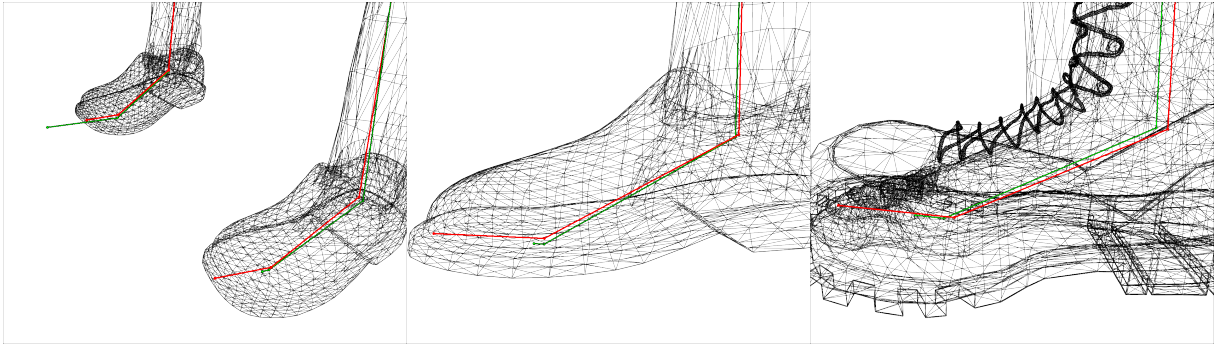
Slika 5.15: Vizualizacija najgorih slučajeva za greške ključnih tačaka kukova.



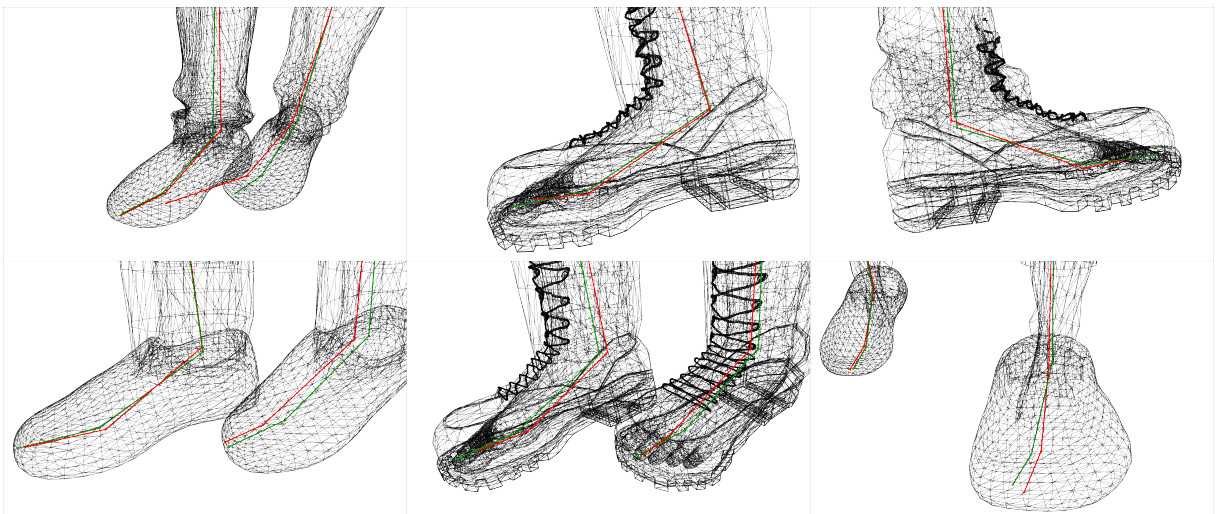
Slika 5.16: Vizualizacija najgorih slučajeva za greške ključnih tačaka koljena.

Na Slici 5.17 prikazani su slučajevi koji proističu kao posljedica dva različita propusta u algoritmu za skraćivanje kostiju koje izlaze izvan 3D modela, s obzirom na strukturu 3D modela generisanih MakeHuman alatom. Pri skraćivanju prednje kosti stopala, zraka emitovana od početka kosti ka njenom kraju, koji bi trebalo da bude pri prednjem kraju stopala, pogađa površinu stopala prije toga. S obzirom na to, ta kost se drastično skraćuje. Ovaj problem je prisutan u 3,8% skupa podataka (11 modela u testnom skupu). Ovo može biti posljedica toga što je obuća predstavljena zasebnim poligonskim mrežama, što nije slučaj kod 3D modela koji nastaju 3D skeniranjem. Vidljivo je da su procijenjene lokacije bolje u odnosu na pretprocesirane stvarne, što donekle ukazuje na robusnost neuronske mreže s obzirom na nepravilno anotirane podatke. Takođe, vidljivo je, na prvom primjeru, da stvarna kost prednjeg dijela desnog stopala nije skraćena. Ovo je posljedica toga što na nekim 3D modelima početni zglob prednjeg dijela stopala ispada izvan 3D modela, te sama zraka koja se emituje od tog zgloba, pri skraćivanju kostiju, ne pogađa površinu 3D modela. Ovaj problem je prisutan na dvije kosti u čitavom skupu podataka.

Na Slici 5.18 prikazani su preostali primjeri slučajeva sa najvećim greškama za ključne tačke stopala. Prvi primjer ilustruje slučaj u kom se ključna tačka dodijeli pogrešnom stopalu. Na preostalim primjerima greške su u prihvatljivim opsezima.



Slika 5.17: Vizualizacija jednog najgoreg slučaja za tačke stopala koji proističe iz propusta u algoritmu za pretprocesiranje podataka.



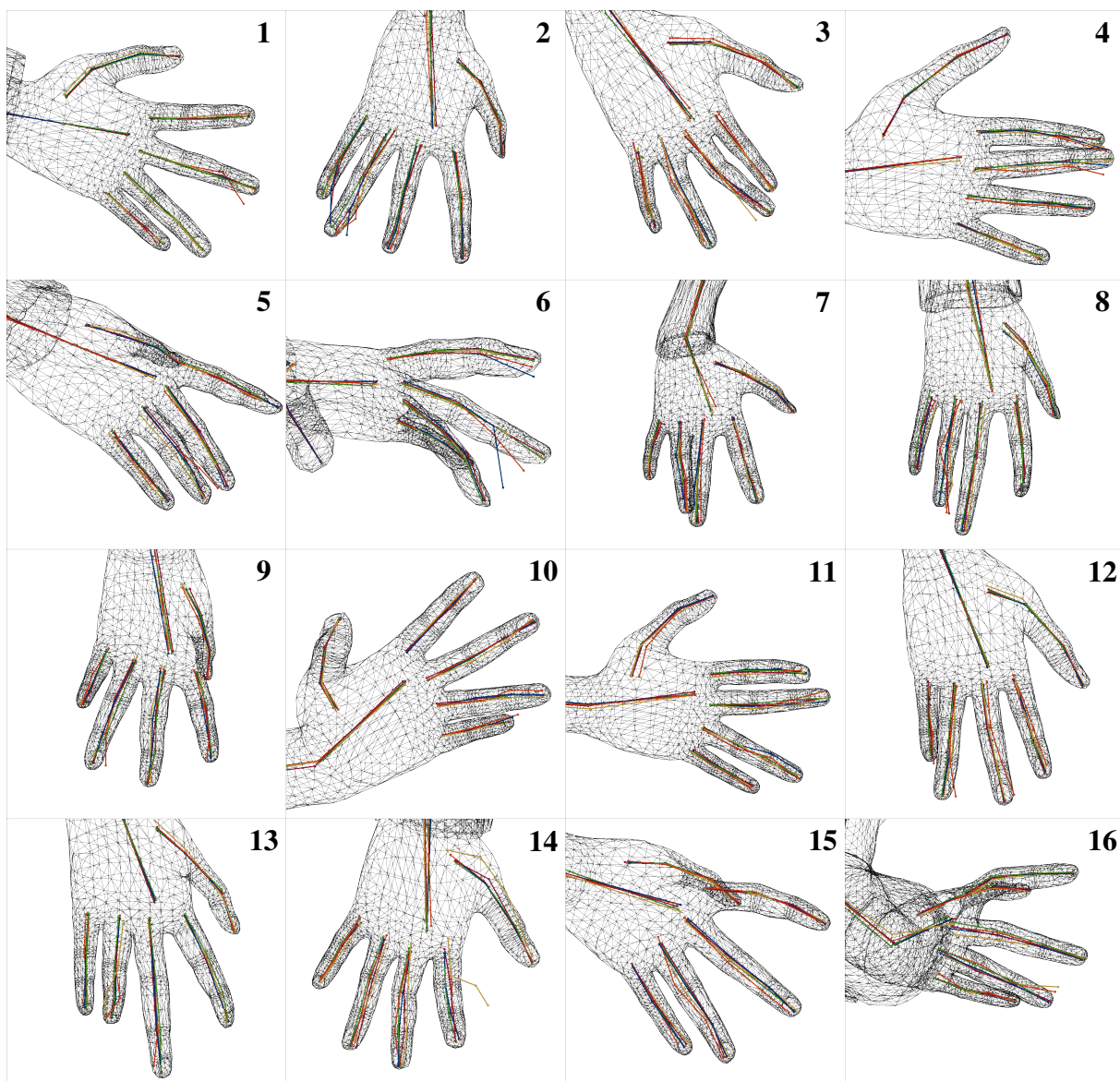
Slika 5.18: Vizualizacija nekih najgorih slučajeva za greške ključnih tačaka stopala.

5.5 Vizualna analiza najgorih slučajeva za tačke prstiju

Na Slici 5.19 prikazani su slučajevi sa najvećim greškama za tačke prstiju, za eksperiment E(LN-DN). Ovi slučajevi ujedno obuhvataju i najgore slučajeve za eksperiment E(LN+SGP-DN). Na Slici su predstavljene stvarne kosti, zelenom bojom, kao i estimacije sljedećih mreža: E(LN-DN), crvenom bojom, E(LN+SGP-DN), plavom bojom, E(LN), žutom bojom i, ljubičastom bojom, E(LN+SGP).

Uočljivo je da postoje značajne greške u lokalizaciji za ove primjere, što uključuje slučajeve gdje su lokalizovane ključne tačke izvan 3D modela, između prstiju, kao i slučajeve u kojima se ključna tačka nalazi unutar pogrešnog prsta. S obzirom na korišteni mehanizam pažnje, ovo znači da se dovoljno značajna pažnja za takve ključne tačke dodjeljuje površinskim tačkama pogrešnih prstiju. Može se uočiti da, za svaku uočenu konkretnu grešku važi da se ona ne manifestuje upotrebom bar jedne od mreža.

Razmatrajući pojedinačne slučajeve prema njihovom rednom broju, može se uočiti sljedeće:



Slika 5.19: Najgori slučajevi za tačke prstiju. Stvarne kosti su obojene zeleno, predikcije E(LN-DN) mreže su obojene u crveno, predikcije E(LN+SGP-DN) varijante su obojene u plavo, predikcije E(LN) mreže su obojene u žuto i predikcije E(LN+SGP) mreže su obojene rozom bojom.

1. E(LN-DN) mreža ima značajnu grešku za tačke srednjeg prsta, gdje je uočljivo ugaono odstupanje i ispadanje tačke na vrhu izvan 3D modela, dok ostale mreže proizvode dobre rezultate,
2. E(LN-DN) i E(LN+SGP-DN) imaju značajne greške za tačke malog i domalog prsta, gdje tačke ispadaju izvan 3D modela, a E(LN+SGP) i E(LN+SGP-DN) pogrešno pozicioniraju tačku za vrh malog prsta pri vrhu domalog prsta, dok E(LN) ima minimalnu grešku, s primjetnim ugaonim odstupanjem za tačke domalog prsta,
3. E(LN-DN) ima primjetno ugaono odstupanje za tačke domalog prsta, kao i prekratku kost

- za vrh malog prsta, dok E(LN+SGP) i E(LN+SGP-DN) korektno pozicioniraju tačke,
4. E(LN-DN) i E(LN+SGP-DN) imaju značajne greške za tačke kažiprsta i srednjeg prsta, s tim da i E(LN+SGP-DN), kao i E(LN-DN), pogrešno pozicioniraju tačku za vrh kažiprsta pri vrhu srednjeg prsta, dok E(LN+SGP) korektno pozicionira sve tačke,
 5. E(LN-DN) ima značajnu grešku za tačke domalog prsta, gdje se estimacije za dvije tačke pri vrhu prsta pozicioniraju između domalog i srednjeg prsta, dok E(LN+SGP-DN) ima najbolje estimacije,
 6. E(LN-DN) i E(LN+SGP-DN) imaju značajne greške za tačke kažiprsta i srednjeg prsta, s tim da su one čak veće za E(LN+SGP-DN), dok E(LN) i E(LN+SGP) imaju najbolje estimacije,
 7. E(LN-DN) i E(LN+SGP-DN) imaju manja ugaona odstupanja za tačke domalog i srednjeg prsta, dok E(LN+SGP) proizvodi najbolje estimacije,
 8. E(LN-DN) ima značajne greške za tačke domalog prsta, sa prisutnim ugaonim odstupanjem i ispadanjem tačaka izvan prsta, što je čak prisutno i za E(LN+SGP) estimacije, dok E(LN+SGP-DN) ima najbolje estimacije,
 9. E(LN-DN) ima ugaona odstupanja za tačke malog prsta i kažiprsta, kao i ispadanje tačke vrha domalog prsta izvan 3D modela, dok E(LN+SGP) i E(LN+SGP-DN) imaju najbolje estimacije,
 10. E(LN-DN) imao primjetne greške za dvije tačke pri vrhu domalog prsta i ispadanje tačke za vrh malog prsta izvan 3D modela, E(LN+SGP-DN) ima primjetno odstupanje za vrh malog prsta, dok E(LN+SGP) ima najbolje estimacije,
 11. E(LN-DN) i E(LN+SGP-DN) imaju primjetna ugaona odstupanja za tačke domalog i srednjeg prsta, dok E(LN+SGP) ima najbolje estimacije,
 12. E(LN-DN) ima značajne greške za tačke malog prsta, domalog prsta i srednjeg prsta, sa ispadanjima tačaka izvan 3D modela, dok E(LN+SGP) i E(LN+SGP-DN) imaju najbolje estimacije,
 13. E(LN-DN) ima značajna ugaona odstupanja za tačke domalog prsta i srednjeg prsta, dok E(LN+SGP) i E(LN+SGP-DN) imaju najbolje estimacije,
 14. E(LN-DN) ima sitna ugaona odstupanja za kosti domalog prsta, dok E(LN+SGP-DN) ima najbolje estimacije,
 15. E(LN-DN) ima blago ugaono odstupanje za tačke malog i srednjeg prsta, kao i značajnije ugaono odstupanje za tačke domalog prsta, dok E(LN+SGP-DN) ima najbolje estimacije,
 16. E(LN-DN) ima grešku u vidu ispadanja tačke na vrhu domalog prsta izvan 3D modela, dok E(LN+SGP) i E(LN+SGP-DN) imaju najbolje estimacije.

Glava 6

Zaključak i budući rad

U ovom radu razmatran je problem konstrukcije algoritma, zasnovanog na dubokom učenju, za lokalizaciju ključnih tačaka tijela i prstiju, predefinisanih tipova, za animacijski skelet 3D modela ljudskog tijela. Opisani su savremeni pristupi rješavanju ovog problema, uključujući modul za šablonske zglobove TARig arhitekture, koji je reimplementiran kako bi se uporedio sa predloženim rješenjem u ovom radu. Uzimajući u obzir oskudicu stvarnih anotiranih 3D skenova, za svrhu poređenja predloženog rješenja sa postojećim rješenjima, kao i međusobno poređenje različitih varijanti predloženog rješenja, formiran je skup sintetičkih 3D modela ljudskog tijela upotrebom MakeHuman alata. Predložene su evaluacione metrike koje su korisne za poređenje različitih modela za rješavanje razmatranog problema.

Prema predloženim evaluacionim metrikama, postignuti su najbolji rezultati u istovremenoj lokalizaciji ključnih tačaka tijela i prstiju za animacijski skelet 3D modela ljudskog tijela. Obezbijeđeno je efikasno rješenje koje zahtijeva minimalnu količinu pretprocesiranja podataka. Predloženi metod za generisanje sintetičkih podataka sa varirajućim oblicima i pozama se može upotrebljavati u daljnjem istraživanju, u svrhu poređenja rješenja za lokalizaciju ključnih tačaka. Predloženo rješenje se može prilagoditi za druge skupove podataka, uključujući stvarne 3D skenove, čineći ga korisnim za smanjivanje količine neophodnog ručnog rada pri formiranju animacijskog skeleta za 3D skenove ljudskog tijela. Ovo je korisno za tokove obrade 3D skenova u svrhu izradu digitalnih avatara i figurina, kao što je tok obrade opisan u [64]. Takođe, predložena arhitektura se može, u apstraktnijem razmatranju, koristiti za lokalizaciju ključnih tačaka bilo kakvih 3D modela, dok god važi pretpostavka da one treba da budu unutar konveksne ljuske modela.

Za predloženu evaluacionu metriku kojom se mjeri procenat tačno lokalizovanih ključnih tačaka, uočeni su i izloženi potencijalni problemi koji mogu nastati primjenom algoritama za određivanje praga udaljenosti za prihvatanje procjene lokacije ključne tačke, što je neophodan korak za izračunavanje procenta tačnosti. Pokazano je da su estimacije za tačke tijela pouzdane,

čak i u najgorim slučajevim. Takođe, pokazano je i da su estimacije za tačke prstiju dobre u većini slučajeva, s tim da, u svakom od eksperimentalnih modela, postoje značajna odstupanja u izuzetnim slučajevima sa najvećom izmjerenom greškom, ostavljajući prostor za buduće istraživanje koje bi se fokusiralo na eliminaciju ovakvih izuzetaka.

Pokazano je i da odabir adekvatnog sloja za normalizaciju može imati značajan uticaj na rezultate, te da arhitekturne izmjene koje proizvode lošije rezultate sa neadekvatnim normalizacionim slojem mogu proizvesti bolje ili uporedive rezultate sa adekvatnijim odabirom normalizacionog sloja. Konkretno, korištenje BatchNorm normalizacije se pokazalo kao neadekvatno, što je čest slučaj ukoliko je *batch* veličina mala. Rezultati sa LayerNorm su znatno bolji, po cijenu vremena obučavanja.

Istraživački kod za predloženo rješenje, koji uključuje implementaciju predložene arhitekture, kao i reimplementaciju TARig–TJM modula, objavljen je u repozitorijumu koda [65].

Po pitanju budućeg rada, predlaže se ispitivanje robusnosti predloženih mreža usljed većih varijacija poze i oblika tijela, kao i istraživanje mogućnosti apliciranja varijacije poze u toku obučavanja, kao vid augmentacije podataka. Predlaže se i istraživanje mogućnosti razvoja arhitekture koja omogućava iterativno poboljšavanje rezultata, s fokusom na tačke prstiju. Takođe, za svrhu lokalizacije ključnih tačaka ljudskog tijela nad stvarnim 3D skenovima, predlaže se pristup zasnovan na principu transfera znanja, u kom se neuronska mreža predtrenira na sintetičkim podacima prije finog podešavanja na stvarnom skupu 3D skenova ljudskog tijela.

LITERATURA

- [1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll i M. J. Black, „SMPL: A Skinned Multi-Person Linear Model,” *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, sv. 34, br. 6, 248:1–248:16, okt. 2015.
- [2] J. Romero, D. Tzionas i M. J. Black, „Embodied Hands: Modeling and Capturing Hands and Bodies Together,” *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 245:1–245:17, sv. 36, br. 6, nov. 2017.
- [3] G. Pavlakos, V. Choutas, N. Ghorbani i dr., „Expressive Body Capture: 3D Hands, Face, and Body from a Single Image,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019., str. 10 975–10 985.
- [4] A. A. A. Osman, T. Bolkart i M. J. Black, „STAR: A Sparse Trained Articulated Human Body Regressor,” u *European Conference on Computer Vision (ECCV)*, 2020., str. 598–613.
- [5] H. Xu, E. G. Bazavan, A. Zanfir, B. Freeman, R. Sukthankar i C. Sminchisescu, „GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020., str. 6184–6193.
- [6] A. A. A. Osman, T. Bolkart, D. Tzionas i M. J. Black, „SUPR: A Sparse Unified Part-Based Human Body Model,” u *European Conference on Computer Vision (ECCV)*, 2022.
- [7] „MakeHuman.” (2024.), adresa: <https://github.com/makehumancommunity/makehuman> (posjećeno 9. 3. 2024.).
- [8] Z. Fang, L. Cai i G. Wang, „MetaHuman Creator: The starting point of the metaverse,” u *2021 International Symposium on Computer Technology and Information Science (ISCTIS)*, 2021., str. 154–157.
- [9] K. Robinette, H. Daanen i E. Paquet, „The CAESAR project: a 3-D surface anthropometry survey,” u *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062)*, 1999., str. 380–386.
- [10] C. Ionescu, D. Papava, V. Olaru i C. Sminchisescu, „Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, sv. 36, br. 7, str. 1325–1339, 2014.

- [11] F. Bogo, J. Romero, M. Loper i M. J. Black, „FAUST: Dataset and Evaluation for 3D Mesh Registration,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014., str. 3794–3801.
- [12] G. Pons-Moll, J. Romero, N. Mahmood i M. J. Black, „Dyna: A Model of Dynamic Human Shape in Motion,” *ACM Transactions on Graphics, (Proc. SIGGRAPH)*, sv. 34, br. 4, 120:1–120:14, avg. 2015.
- [13] C. Zhang, S. Pujades, M. Black i G. Pons-Moll, „Detailed, Accurate, Human Shape Estimation from Clothed 3D Scan Sequences,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, jul 2017., str. 5484–5493.
- [14] A. Saint, E. Ahmed, A. E. R. Shabayek i dr., „3DBodyTex: Textured 3D Body Dataset,” u *International Conference on 3D Vision (3DV)*, 2018., str. 495–504.
- [15] S. Yan, J. Wirta i J.-K. Kämäräinen, „Anthropometric clothing measurements from 3D body scans,” *Machine Vision and Applications*, sv. 31, 2019.
- [16] S. Yan, J. Wirta i J.-K. Kämäräinen, „Silhouette Body Measurement Benchmarks,” u *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021., str. 7804–7809.
- [17] Z. Zheng, T. Yu, Y. Wei, Q. Dai i Y. Liu, „DeepHuman: 3D Human Reconstruction from a Single Image,” u *The IEEE International Conference on Computer Vision (ICCV)*, okt. 2019.
- [18] G. Tiwari, B. L. Bhatnagar, T. Tung i G. Pons-Moll, „SIZER: A Dataset and Model for Parsing 3D Clothing and Learning Size Sensitive 3D Clothing,” u *European Conference on Computer Vision (ECCV)*, Springer, avg. 2020.
- [19] T. Yu, Z. Zheng, K. Guo, P. Liu, Q. Dai i Y. Liu, „Function4D: Real-time Human Volumetric Capture from Very Sparse Consumer RGBD Sensors,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2021.
- [20] S. S. Jinka, R. Chacko, A. Srivastava, A. Sharma i P. J. Narayanan, „SHARP: Shape-Aware Reconstruction of People in Loose Clothing,” *International Journal of Computer Vision*, sv. 131, str. 918–937, 2022.
- [21] X. Zhu, T. Liao, X. Zhang i dr., „MVP-Human Dataset for 3-D Clothed Human Avatar Reconstruction From Multiple Frames,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, sv. 5, br. 4, str. 464–475, 2023.
- [22] Z. Su, T. Yu, Y. Wang i Y. Liu, „DeepCloth: Neural Garment Representation for Shape and Style Editing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, sv. 45, br. 2, str. 1581–1593, 2023.

- [23] *ISO/IEC 19774-1:2019 Humanoid Animation (H-Anim) – Part 1: Architecture*, International Organization for Standardization, ISO/IEC JTC 1/SC 24, 2019. adresa: <https://www.iso.org/standard/74163.html>.
- [24] *ISO/IEC 19774-2:2019 Humanoid Animation (H-Anim) – Part 2: Motion capture data*, International Organization for Standardization, ISO/IEC JTC 1/SC 24, 2019. adresa: <https://www.iso.org/standard/74164.html>.
- [25] R. Parent, *Computer Animation: Algorithms and Techniques*, 3. izdanje. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2012., ISBN: 9780124158429.
- [26] T. O’Hailey, *Rig it Right! Maya Animation Rigging Concepts*, 3rd. USA: CRC Press, 2024., ISBN: 9781032555232.
- [27] Y. Yang, „The skinning in character animation: A survey,” *Academic Journal of Computing & Information Science*, sv. 5, br. 4, str. 4–17, 2022.
- [28] R. Q. Charles, H. Su, M. Kaichun i L. J. Guibas, „PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017., str. 77–85.
- [29] C. R. Qi, L. Yi, H. Su i L. J. Guibas, „PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space,” u *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS’17, Long Beach, California, USA: Curran Associates Inc., 2017., str. 5105–5114, ISBN: 9781510860964.
- [30] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein i J. M. Solomon, „Dynamic Graph CNN for Learning on Point Clouds,” *ACM Transactions on Graphics (TOG)*, 2019.
- [31] G. Qian, Y. Li, H. Peng i dr., „PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies,” u *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [32] H. Lin, X. Zheng, L. Li i dr., „Meta Architecture for Point Cloud Analysis,” u *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, jun 2023., str. 17 682–17 691.
- [33] Z. Xu, Y. Zhou, E. Kalogerakis, C. Landreth i K. Singh, „RigNet: Neural Rigging for Articulated Characters,” *ACM Transactions on Graphics*, sv. 39, br. 4, 58:1–58:14, avg. 2020.
- [34] J. Ma i D. Zhang, „TARig: Adaptive template-aware neural rigging for humanoid characters,” *Computers & Graphics*, sv. 114, str. 158–167, 2023., ISSN: 0097-8493.
- [35] Z. Weng, A. S. Gorban, J. Ji, M. Najibi, Y. Zhou i D. Anguelov, „3D Human Keypoints Estimation from Point Clouds in the Wild without Human Labels,” u *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, jun 2023., str. 1158–1167.

- [36] L. Ge, Y. Cai, J. Weng i J. Yuan, „Hand PointNet: 3D Hand Pose Estimation Using Point Sets,” u *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018., str. 8417–8426.
- [37] L. Ge, Z. Ren i J. Yuan, „Point-to-Point Regression PointNet for 3D Hand Pose Estimation,” u *Proceedings of the European Conference on Computer Vision (ECCV)*, sept. 2018.
- [38] X. Chen, G. Wang, C. Zhang, T.-K. Kim i X. Ji, „SHPR-Net: Deep Semantic Hand Pose Regression From Point Clouds,” *IEEE Access*, sv. 6, str. 43 425–43 439, 2018.
- [39] W. Cheng, J. H. Park i J. H. Ko, „HandFoldingNet: A 3D Hand Pose Estimation Network Using Multiscale-Feature Guided Folding of a 2D Hand Skeleton,” u *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Los Alamitos, CA, USA: IEEE Computer Society, okt. 2021., str. 11 240–11 249.
- [40] S. Novaković i V. Risojević, „Learning Localization of Body and Finger Animation Skeleton Joints on Three-Dimensional Models of Human Bodies,” u *2024 Zooming Innovation in Consumer Technologies Conference (ZINC) (ZINC 2024)*, Novi Sad, Serbia, maj 2024., str. 6.
- [41] W. Rudin, *Principles of Mathematical Analysis* (International series in pure and applied mathematics). McGraw-Hill, 1976., ISBN: 9780070856134.
- [42] T. Tao, *Analysis II: Third Edition* (Texts and Readings in Mathematics). Springer Nature Singapore, 2016., ISBN: 9789811018046.
- [43] D. Huttenlocher, G. Klanderman i W. Rucklidge, „Comparing images using the Hausdorff distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, sv. 15, br. 9, str. 850–863, 1993.
- [44] „Hausdorff distance — Wikipedia, The Free Encyclopedia.” (2024.), adresa: https://en.wikipedia.org/w/index.php?title=Hausdorff_distance&oldid=1213656182%22 (posjećeno 11. 5. 2024.).
- [45] J. T. Henrikson, „Completeness and Total Boundedness of the Hausdorff Metric,” 1999.
- [46] O. Ronneberger, P. Fischer i T. Brox, „U-Net: Convolutional Networks for Biomedical Image Segmentation,” u *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells i A. F. Frangi, urednici, Cham: Springer International Publishing, 2015., str. 234–241.
- [47] K. Yin, H. Huang, D. Cohen-Or i H. Zhang, „P2P-NET: Bidirectional Point Displacement Net for Shape Transform,” *ACM Transactions on Graphics (Special Issue of SIGGRAPH)*, sv. 37, br. 4, 152:1–152:13, 2018.
- [48] Y. Cheng, „Mean shift, mode seeking, and clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, sv. 17, br. 8, str. 790–799, 1995.

- [49] A. Shamir i D. Cohen-Or, „Consistent Mesh Partitioning and Skeletonisation using the Shape Diameter Function,” *The Visual Computer*, sv. 24, str. 249–259, apr. 2008.
- [50] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin i M. Jorge Cardoso, „Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations,” u *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Cham: Springer International Publishing, 2017., str. 240–248, ISBN: 978-3-319-67558-9.
- [51] D. Kingma i J. Ba, „Adam: A Method for Stochastic Optimization,” *International Conference on Learning Representations*, dec. 2014.
- [52] „SMPL-X: A new joint 3D model of the human body, face and hands together, Model parameter transfer.” (2024.), adresa: https://github.com/vchoutas/smplx/tree/main/transfer_model (posjećeno 9. 3. 2024.).
- [53] R. T. Rockafellar, *Convex Analysis*. Princeton: Princeton University Press, 1970., ISBN: 9781400873173.
- [54] R. Rockafellar, M. Wets i R. Wets, *Variational Analysis* (Grundlehren der mathematischen Wissenschaften). Springer Berlin Heidelberg, 2009., ISBN: 9783540627722.
- [55] J. L. Ba, J. R. Kiros i G. E. Hinton, *Layer Normalization*, 2016. arXiv: 1607.06450 [stat.ML].
- [56] A. L. Maas, A. Y. Hannun, A. Y. Ng i dr., „Rectifier Nonlinearities Improve Neural Network Acoustic Models,” u *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, sv. 30, 2013., str. 3.
- [57] „Blender.” (2024.), adresa: <https://www.blender.org> (posjećeno 9. 3. 2024.).
- [58] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli i G. Ranzuglia, „MeshLab: An Open-Source Mesh Processing Tool,” u *Eurographics Italian Chapter Conference*, V. Scarano, R. D. Chiara i U. Erra, urednici, The Eurographics Association, 2008., ISBN: 978-3-905673-68-5.
- [59] M. Garland i P. S. Heckbert, „Surface simplification using quadric error metrics,” u *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '97, USA: ACM Press/Addison-Wesley Publishing Co., 1997., str. 209–216, ISBN: 0897918967.
- [60] „Trimesh.” (2024.), adresa: <https://trimesh.org> (posjećeno 9. 3. 2024.).
- [61] Q.-Y. Zhou, J. Park i V. Koltun, „Open3D: A Modern Library for 3D Data Processing,” *arXiv:1801.09847*, 2018.
- [62] A. Shamir i D. Cohen-Or, „Consistent Mesh Partitioning and Skeletonisation using the Shape Diameter Function,” *The Visual Computer*, sv. 24, str. 249–259, apr. 2008.
- [63] I. Loshchilov i F. Hutter, „Decoupled Weight Decay Regularization,” u *International Conference on Learning Representations*, 2017.

- [64] S. Ni, R. Luo, Y. Zhang i dr., „Scan2Avatar: Automatic Rigging for 3D Raw Human Scans,” u *ACM SIGGRAPH 2020 Posters*, ser. SIGGRAPH '20, Virtual Event, USA: Association for Computing Machinery, 2020., ISBN: 9781450379731.
- [65] „Learning Localization of Body and Finger Animation Skeleton Joints on 3D Models of Human Bodies (research code).” (2024.), adresa: <https://github.com/sznov/joint-localization> (posjećeno 9. 3. 2024.).

Biografija autora

Lični podaci

Rođen 12.3.1996. u Banjoj Luci.

Obrazovanje

Završio srednju Elektrotehničku školu „Nikola Tesla“ u Banjoj Luci.

Prvi ciklus studija upisao 2015. godine na Elektrotehničkom fakultetu Univerziteta u Banjoj Luci, studijski program Računarstvo i informatika, smjer Softversko inženjerstvo. Diplomirao 2020. godine, sa prosječnom ocjenom 8,98. Tema diplomskog rada: „Bočni efekti prekorednog izvršavanja instrukcija u savremenim procesorima“.

Drugi ciklus studija upisao 2020. godine, na Elektrotehničkom fakultetu Univerziteta u Banjoj Luci, studijski program Računarstvo i informatika.

Radno iskustvo

Od 2020. godine zaposlen na Elektrotehničkom fakultetu Univerziteta u Banjoj Luci, kao saradnik u zvanju asistenta. Zadužen na predmetima: Programski jezici 1, Formalne metode u softverskom inženjerstvu, Osnovi operativnih sistema, Arhitektura računara, Odabrana poglavlja iz operativnih sistema, Operativni sistem za rad u realnom vremenu i Računarska grafika.

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ПОДАЦИ О АУТОРУ ОДБРАЊЕНОГ МАСТЕР/МАГИСТАРСКОГ РАДА

Име и презиме аутора мастер/магистарског рада: **Стефан Новаковић**

Датум, мјесто и држава рођења аутора: **12.3.1996, Бања Лука, Босна и Херцеговина**

Назив завршеног факултета/Академије аутора и година дипломирања:
Електротехнички факултет Универзитета у Бањој Луци, 2020. године

Датум одбране завршног/дипломског рада аутора: **28.1.2020.**

Наслов завршног/дипломског рада аутора:
Бочни ефекти прекоредног извршавања инструкција у савременим процесорима

Академско звање коју је аутор стекао одбраном завршног/дипломског рада:
дипломирани инжењер електротехнике (240 ECTS)

Академско звање које је аутор стекао одбраном мастер/магистарског рада:
мастер електротехнике - 300 ECTS - Рачунарство и информатика

Назив факултета/Академије на коме је мастер/магистарски рад одбрањен:
Електротехнички факултет Универзитета у Бањој Луци

Наслов мастер/магистарског рада и датум одбране:
**Локализација кључних тачака тродимензионалних модела људског тијела,
3.7.2024.**

Научна област мастер/магистарског рада према CERIF шифрарнику: **T 111**

Имена ментора и чланова комисије за одбрану мастер/магистарског рада:
проф. др Зденка Бабић, предсједник
проф. др Владимир Рисојевић, ментор
доц. др Михајло Савић, члан

У Бањој Луци, дана 21. 6. 2024.

Декан


ИЗЈАВА О АУТОРСТВУ

Изјављујем да је
мастер/магистарски рад

Наслов рада: Локализација кључних тачака тродимензионалних модела људског тијела

Наслов рада на енглеском језику: Keypoint Localization on Three-Dimensional Models of Human Body

- резултат сопственог истраживачког рада,
- да мастер/магистарски рад, у цјелини или у дијеловима, није био предложен за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

У Бањој Луци 25.6.2024.

Потпис кандидата

Емирхан Новакчић

Изјава којом се овлашћује Електротехнички факултет/ Академија умјетности Универзитета у Бањој Луци да мастер/магистарски рад учини јавно доступним

Овлашћујем Електротехнички факултет/ Академију умјетности Универзитета у Бањој Луци да мој мастер/магистарски рад, под насловом

Локализација кључних тачака тродимензионалних модела људског тијела

који је моје ауторско дјело, учини јавно доступним.

Мастер/магистарски рад са свим прилозима предао/ла сам у електронском формату, погодном за трајно архивирање.

Мој мастер/магистарски рад, похрањен у д и г и т а л н и р е п о з и т о р и ј у м Универзитета у Бањој Луци, могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (*Creative Commons*), за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство - некомерцијално - без прераде
4. Ауторство - некомерцијално - дијелити под истим условима
5. Ауторство - без прераде
6. Ауторство - дијелити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

У Бањој Луци 25.6.2024.

Потпис кандидата

Џељко Коњковић

Изјава о идентичности штампане и електронске верзије
мастер/магистарског рада

Име и презиме аутора: Стефан Новаковић

Наслов рада: Локализација кључних тачака тродимензионалних модела људског тијела

Ментор: проф. др Владимир Рисојевић

Изјављујем да је штампана верзија мог мастер/магистарског рада идентична електронској верзији коју сам предао/ла за дигитални репозиторијум Универзитета у Бањој Луци.

У Бањој Луци 25.6.2024.

Потпис кандидата

Стефан Новаковић

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ
Патре 5
78000 Бања Лука

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ БАЊА ЛУКА	
Број:	454
Датум:	07.06.2024.

Др Зденка Бабић, редовни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Владимир Рисојевић, ванредни професор
Универзитет у Бањој Луци, Електротехнички факултет

Др Михајло Савић, доцент
Универзитет у Бањој Луци, Електротехнички факултет

НАУЧНО – НАСТАВНОМ ВИЈЕЋУ ЕЛЕКТРОТЕХНИЧКОГ ФАКУЛТЕТА УНИВЕРЗИТЕТА У БАЊОЈ ЛУЦИ

Одлуком Научно–наставног вијећа Електротехничког факултета Универзитета у Бањој Луци број 20/3.805-15/23 од 11.09.2023. године, именовани смо за чланове Комисије за завршни рад другог циклуса студија кандидата Стефана Новаковића, дипл. инж. електротехнике, под називом „Локализација кључних тачака тродимензионалних модела људског тијела“. Након прегледа приложеног рада, подносимо сљедећи

ИЗВЈЕШТАЈ

1. БИОГРАФСКИ ПОДАЦИ КАНДИДАТА

Стефан Новаковић је рођен 12.03.1996. године у Бањој Луци. Средњу Електротехничку школу „Никола Тесла“ у Бањој Луци је завршио 2015. године. Исте године је уписао први циклус студија на Електротехничком факултету Универзитета у Бањој Луци, студијски програм Рачунарство и информатика, смјер Софтверско инжењерство. Дипломирао је 2020. године, одбраном дипломског рада: „Бочни ефекти прекоредног извршавања инструкција у савременим процесорима“ и са просјечном оцјеном у току студија 8,98. Други циклус студија је уписао 2020. године, на Електротехничком факултету Универзитета у Бањој Луци, студијски програм Рачунарство и информатика и положио све испите предвиђене планом и програмом. Од 2020. године је запослен на Електротехничком факултету као сарадник у звању асистента на катедри за Рачунарску технику и информатику. Задужен је на предметима: Програмски језици 1, Формалне методе у софтверском инжењерству, Основи оперативних система, Архитектура рачунара, Одабрана поглавља из оперативних система, Оперативни системи за рад у реалном времену и Рачунарска графика. До сада је објавио један рад:

1. S. Novaković, V. Risojević, „Learning Localization of Body and Finger Animation Skeleton Joints on Three-Dimensional Models of Human Bodies,“ у 2024 *Zooming*

2. ОСНОВНИ ПОДАЦИ О РАДУ

Завршни рад другог циклуса студија кандидата Стефана Новаковића, дипл. инж. електротехнике, под називом “Локализација кључних тачака тродимензионалних модела људског тијела” има обим од 70 страница. Садржи насловну страну на српском и енглеском језику, информације о ментору и раду на српском и енглеском језику, садржај, 4 табеле и 35 слика. Рад је организован у седам поглавља:

1. Увод
2. Преглед литературе
3. Архитектура рјешења
4. Скуп података и евалуационе метрике
5. Експериментални резултати
6. Закључак и будући рад

На крају рада дати су преглед кориштене литературе (64 референце), CD са изворним кодом и биографија кандидата.

У првој глави завршног рада је дефинисан проблем локализације кључних тачака тродимензионалних модела људског тијела и изложена мотивација за његово рјешавање. Такође је дата таксономија постојећих приступа за локализацију кључних тачака са посебним нагласком на примјену дубоког учења. У овој глави је изложена и структура самог рада и наведени доприноси рада.

У другој глави је дат преглед литературе у областима анализе скупова тачака и естимације анимацијског скелета. Формално су описане приступи засновани на неуронским мрежама који су показали најбоље резултате у поменутих областима. Глава се завршава прегледом расположивих скупова података и параметарских модела људског тијела са детаљним описом алата MakeHuman који је у овом раду кориштен за генерисање синтетичких података за обучавање и тестирање.

У трећој глави је дат приједлог архитектуре рјешења која израчунава локације кључних тачака као конвексне комбинације позиција улазних тачака. Коефицијенти конвексне комбинације естимирају се помоћу граф-конволуционе неуронске мреже обучене тако да се минимизира функција цијене израчуната као сума квадрата Еуклидових удаљености између локација стварних кључних тачака и процијењених локација кључних тачака.

Материјал и методологија експеримената извршених у оквиру практичног дијела рада су описани у четвртој глави. За обучавање и тестирање предложеног рјешења генерисан је синтетички скуп података. Конкретно, кориштен је Mass Produce модул MakeHuman софтверског алата за генерисање 3D модела људског тијела при чему су варирани параметри позе, облика и одјеће генерисаних модела. Сви генерисани модели садрже и анимацијски скелет са означеним кључним тачкама. За оцјену перформанси локализације кључних тачака кориштене су двије метрике: просјечна грешка позиције по кључној тачки и проценат коректних кључних тачака.

Експериментални резултати су приказани у петој глави. Обучено је и тестирано осам неуронских мрежа које садрже различите варијације у архитектури и начину обраде улазних података. Такође, у циљу поређења резултата са актуелним стањем у области, реимплементирана је и обучена једна неуронска мрежа из литературе. Поређењем резултата уочено је да су сви предложени методи бољи у односу на реимплементацију рјешења из литературе. Анализиран је и утицај различитих варијација основне мреже на перформансе локализације кључних тачака као и са становишта времена извршавања, броја параметара мреже и заузећа меморије. Визуелизовани су карактеристични примјери детекције кључних тачака и улазних тачака које највише утичу на те детекције што доприноси објашњивости метода за локализацију. Такође, урађена је детаљна квантитативна и квалитативна анализа грешака локализације на тестном скупу.

У шестој глави су, на основу експерименталних резултата, изведени закључци. Изложене су главне предности и недостаци имплементираних метода за локализацију кључних тачака тродимензионалног модела људског тијела, те потенцијални правци за даље истраживање у предметној области.

3. АНАЛИЗА И НАЈВАЖНИЈИ ДОПРИНОСИ РАДА

Разматрајући завршни рад кандидата Стефана Новаковића, Комисија је закључила да својим садржајем, постигнутим резултатима и закључцима задовољава критеријуме који се постављају пред завршни рад другог циклуса студија. Рад у цјелини има добро систематизовану структуру и план излагања. Наслов рада је јасно формулисан, разумљив, прецизно описује предмет истраживања и у потпуности указује на садржај рада.

Свеобухватном теоријском анализом као и конкретним експерименталним радом, кандидат Стефан Новаковић је показао зрелост и способност да савлада и систематизује знања из једне истраживачке области.

Имајући у виду значај проблема локализације кључних тачака тродимензионалних модела људског тијела, те актуелност истраживања техника анимације људског скелета, овај завршни рад обухвата актуелна истраживања и представља допринос стању у области.

Комисија констатује да је рад написан у складу са образложењем у пријави теме, као и да су остварени сви резултати који су били и планирани у образложењу пријаве теме:

А. Систематизација савремених метода за локализацију кључних тачака 3D модела људског тијела и шака

У раду је дат преглед литературе из области локализације кључних тачака тродимензионалних модела људског тијела. Идентификоване су и описане основне групе приступа који се користе за рјешавање овог проблема и размотрене њихове добре и лоше стране. Посебна пажња је посвећена кориштењу дубоког учења над тродимензионалним моделима. Дат је формални, математички, опис метода за анализу скупова тачака, као и метода за естимацију анимацијског скелета који користе дубоко учење. Такође, анализирани су постојећи скупови података који се у литератури користе за рјешавање овог и сродних проблема. Уочен је недостатак квалитетних отворених скупова података који би омогућили обучавање неуронских мрежа за локализацију кључних тачака.

Б. Приједлог метода за генерисање синтетичког скупа 3D модела људског тијела са анимацијским скелетом у варирајућим позама и генерисање таквог скупа,

Имајући у виду уочени недостатак адекватних скупова података, предложено је кориштење софтвера за генерисање фотореалистичних 3D модела људских тијела на

основу конфигурационих параметара. На овај начин су генерисани скупови за обучавање, валидацију и тестирање неуронске мреже који садрже моделе људског тијела са варијацијама позе, облика и одјеће.

В. Конструкција архитектуре неуронске мреже за локализацију кључних тачака 3D модела људског тијела, обучавање и евалуација конкретне неуронске мреже на основу конструисане архитектуре и документовање утицаја хиперпараметара.

Предложен је и имплементиран метод за локализацију кључних тачака заснован на кориштењу граф-конволуционих неуронских мрежа. Предложени метод подржава и прецизну естимацију кључних тачака шака. Обучена неуронска мрежа у фази тестирања даје резултате који надмашују резултате добијене приступима из литературе, према адекватним евалуационим метрикама. Испитан је утицај различитих варијација архитектуре мреже и хиперпараметара на постигнуте перформансе. Од посебног је значаја детаљна квантитативна и квалитативна анализа грешака локализације које мрежа прави на тестном скупу.

Дио резултата истраживања проведених у оквиру овог завршног рада је приказан у раду:

S. Novaković, V. Risojević, „Learning Localization of Body and Finger Animation Skeleton Joints on Three-Dimensional Models of Human Bodies,” u *2024 Zooming Innovation in Consumer Technologies Conference (ZINC) (ZINC 2024)*, Novi Sad, Srbija, мај 2024.

4. ЗАКЉУЧАК И ПРИЈЕДЛОГ

Комисија сматра да завршни рад другог циклуса студија под називом „Локализација кључних тачака тродимензионалних модела људског тијела“, кандидата Стефана Новаковића, дипл. инж. електротехнике, садржи све потребне елементе и резултате којима су остварени постављени циљеви истраживања, те предлаже Научно-наставном вијећу Електротехничког факултета Универзитета у Бањој Луци да усвоји извјештај Комисије и одобри заказивање јавне усмене одбране.


Бања Лука, 06.06.2024. године

Комисија:

Проф. др Зденка Бабић, предсједник



Проф. др Владимир Рисојевић, ментор



Доц. др Михајло Савић, члан