

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ STUDIJŲ PROGRAMA

**Kompiuterinės regos taikymas sporto ir
reabilitacijos informacinėje sistemoje**

**Application of Computer Vision in Sport and Rehabilitation
Information System**

Magistro baigiamasis darbas

Atliko: Ernest Jaščanin

Darbo vadovas: doc. dr. Linas Petkevičius

Recenzentas: j.asist. Boleslovas Dapkūnas

Vilnius – 2024

Dėkoju darbo vadovui Linui Petkevičiui už idėjas ir nuolatinę pagalbą.
Dėkoju Vytautui Valaičiui už idėjas ir dalyvavimą duomenų rinkinio filmavime.
Nuoširdžiai dėkoju Tėvams, Seneliams bei artimiems draugams už paramą, supratingumą ir
begalinę kantrybę šiame nelengvame kelyje.

Santrauka

Šiame darbe yra tiriamas Tinetti pereinamųjų pratimų identifikavimas ir vertinimas, pasinaudojus žmogaus pozos taškų pagrindu veikiančiais giliaisiais neuroniniais tinklais. Mokslinės literatūros analizės metu nustatyti egzistuojantys duomenų rinkiniai nėra tinkami korektiškumo nustatymui, todėl yra naujai sukurtas Tinetti duomenų rinkinys, turintis teisingų ir neteisingų pratimų atlikimo pavyzdžių. Surinktas duomenų rinkinys yra praplėstas duomenimis iš kelių egzistuojančių duomenų rinkinių. Pasinaudojus MediaPipe BlazePose yra gauta žmogaus pozos taškų informacija. Tyrimo metu nustatyti ir pasiūlyti daugiatislio konvoliucinio neuroninio tinklo architektūros pavyzdžiai gebantys nustatyti Tinetti pratimą ir jo korektiškumą. Geriausias pasiūlytas modelis pasiekia 77.65% pratimo klasifikavimo tikslumą ir 65.7% pratimo teisingumo tikslumą.

Raktiniai žodžiai: Mašininis mokymasis, Gilusis mokymasis, Kompiuterinė rega, Reabilitacijos Tinetti duomenų rinkinys, Daugiatislio uždavinio klasifikavimas

Summary

In this Master thesis, the identification and evaluation of Tinetti transition exercises using human pose based deep neural networks is being analyzed. The identified existing datasets are not suitable for correctness estimation, therefore a new Tinetti dataset is created containing examples of performing correct and incorrect exercise versions. The collected dataset is augmented with data from several existing datasets. Human pose keypoints are extracted using MediaPipe BlazePose. Identified and proposed the examples of multi-objective CNN neural network architectures capable of determining the Tinetti exercise and its correctness. The best suggested model achieved 77.65% exercise classification accuracy and 65.7% exercise correctness accuracy.

Keywords: Machine learning, Deep learning, Computer vision, Rehabilitation Tinetti dataset, Multitask classification

Turinys

ĮVADAS	7
1. GILIEJI NEURONINIAI TINKLAI	11
1.1. Konvoliuciniai neuroniniai tinklai	11
1.2. Rekurentiniai neuroniniai tinklai	18
2. VEIKSMO ATPAŽINIMO SISTEMOS	21
2.1. Veiksmo atpažinimo sistemų problemos	21
2.2. Požymių nustatymo būdai	23
2.3. Skeleto pagrindu žmogaus pozos nustatymo sistema	24
2.4. Skeleto pagrindu veiksmo nustatymo sistemos	25
2.5. Pratimų ir susijusių sistemų apžvalga	28
2.6. Daugiatikslio uždavinio optimizavimas	30
3. VEIKSMŲ ATPAŽINIMO DUOMENŲ RINKINIAI	31
4. TINETTI DUOMENŲ RINKINYS	33
4.1. Sukurtas duomenų rinkinys	33
4.2. Duomenų rinkinio praplėtimas	35
5. EKSPERIMENTAI	36
5.1. Duomenų paruošimas	36
5.2. Tyrimo aplinka	36
5.3. Modelio pasirinkimas	36
5.4. Modelių architektūros	37
5.4.1. M2	37
5.4.2. M3	37
5.4.3. M4	38
5.4.4. M5	38
5.4.5. Kiti modeliai	38
5.5. Mokymo epizodų pasirinkimas	38
5.6. Klasifikavimo slenkstinių reikšmių paieška	38
5.7. Dviejų klasių klasifikavimo eksperimentai	40
5.7.1. Tikslas	40
5.7.2. Konfigūracija	40
5.7.3. Rezultatai	40
5.8. Daugelio klasių klasifikavimo eksperimentai	44
5.8.1. Tikslas	44
5.8.2. Konfigūracija	44
5.8.3. Rezultatai	44
5.9. Daugiatikslio klasifikavimo eksperimentai	45
5.9.1. Tikslas	45
5.9.2. Modeliai	45
5.9.2.1. Multi1	45
5.9.2.2. Multi2	45
5.9.2.3. Multi3	45
5.9.2.4. Multi4	45
5.9.2.5. Multi5	46
5.9.3. Konfigūracija	46
5.9.4. Rezultatai	46

5.10. Agreguotų kadru eksperimentai	49
5.11. Neteisingų atvejų testavimas	50
5.12. Daugiatikslio klasifikavimo optimalių sprendinių paieška	51
REZULTATAI IR IŠVADOS	54
ŠALTINIAI	56
PRIEDAI	60
1 priedas. TINETTI Testas	60
2 priedas. Dviejų/daugelio klasių klasifikavimo modeliai	62
3 priedas. Daugiatikslio klasifikavimo modeliai	65

Įvadas

Pastaraisiais metais kompiuterinės regos technologija yra aktyviai tyrinėjama mokslinės bendruomenės. Mokslininkai iš įvairių šalių bando atrasti jos pritaikymą skirtingose gyvenimo srityse.

Viena iš šių sričių, kurioje kompiuterinė rega gali turėti didelį poveikį, yra medicina. Įvairių tipų medicinos vaizdų analizė, Parkinsono ligos diagnozavimo [CCL⁺09] iki skirtingo tipo reabilitacijų taikymo, pavyzdžiui, žmonių patyrusių insultą, pirštų reabilitacijos [QSS⁺18; VDD15]. Apmokyto personalo darbai su reabilitacijos pacientais trūkumas [VDD15] sąlygoja tolimesnę technologijos vystymąsi. Nepaisant gaunamos naudos, nustatytos atliktame tyrime [Jun18], jau atrasti sprendimai nėra masiškai įdiegiami Lietuvos ligoninėse ar reabilitacijos centruose dėl sprendimų aukštos kainos.

Kita galima pritaikymo sritis – sportas ir reabilitacija. Sportininkai yra suinteresuoti stebėti savo veiksmus varžybų ir treniruočių metu [STA12]. Šiam tikslui pasiekti galima pasinaudoti kompiuterinės regos suteikiamomis galimybėmis sekti rezultatus realiu laiku, juos išsaugoti bei vėliau analizuoti klaidas. Dėka to, įmanoma sukurti specialius sportininkų profilius ir pritaikyti asmeninę treniruočių programą [BB08]. Visgi realaus laiko samprata priklauso nuo naudojamo konteksto, pavyzdžiui, panaudojus vieną iš egzistuojančių sprendimų galima pasiekti 24 kadrų per sekundę HD raiškos paveikslėlių apdorojimo greitį [PBK⁺12]. Vertinant galimą panaudojimą reabilitacijoje, svarbu suprasti kokie tikslai yra keliami toje srityje. Reabilitacijos tikslas yra padėti žmogui atstatyti pilnai arba dalinai prarastas, arba pagerinti esamus gebėjimus, reikalingus kasdieniam gyvenimui. Gebėjimų praradimas gali atsirasti dėl įvairių lūžių, amputacijų, insulto ar kito neurologinio sutrikimo, artrito, širdies veiklos sutrikimo ar kitų medicininių priežasčių [Isa23]. Priklausomai nuo situacijos, į reabilitacijos planą gali būti įtrauktos tokios veiklos:

- Kognityvinė reabilitacijos terapija, padėsianti persimokyti arba patobulinti tokius įgūdžius kaip mąstymas, mokymasis, atmintį, planavimą ir sprendimų priėmimą;
- Konsultavimas psichikos sveikatos klausimais;
- Muzikos arba meno terapija, padėsianti išmokti išreikšti jausmus ir sukurti socialinį ratą;
- Konsultavimas mitybos klausimais;
- Skausmo gydymas;
- Ergoterapija, padėsianti atlikti kasdienes darbus;
- Kineziterapija, padėsianti sustiprinti jėgą, išlavinti judrumą, ir kūno pasirengimą fizinei veiklai;
- Pramoginė, kalbos ir kitos terapijos.

Kompiuterinės regos technologija evoliucionavo nuo rankinio skaitmeninimo sistemų, automatinų žymeklio sistemų iki be žymeklio judesio analizės sistemų. Pastarosios leidžia geriau analizuoti žmonių judesius realiomis sąlygomis, kadangi nėra reikalaujama turėti papildomų prietaisų ant žmogaus kūno [CEC⁺18].

Be žymeklio sistema yra sudaryta iš keturių pagrindinių komponentų: kamerų sistemos, žmogaus kūno modelio, paveikslėlių požymių (angl. features) ir algoritmų, skirtų identifikuoti kūno

modelio parametrus. Ši sistema veikia išankstinio duomenų rinkimo principu, kurie naudojami arba modelio kūrimui, arba skiriamąjo algoritmo apmokymui neuroninių tinklų pagalba [CEC⁺18].

Algoritmai, nustatantys kūno pozą, skirstomi į generuojančius (angl. generative) ir skiriamuosius (angl. discriminative). Generuojantys algoritmai kuria hipotezę pagal modelio parametrus ir palygina su paveikslėlio duomenimis, vėliau hipotezė yra iteratyviai tikslinama, kol surandama geriausia įmanoma atitiktis. Tuo tarpu skiriamieji algoritmai tiesiogiai lygina paveikslėlio duomenis su modelio parametrais [CEC⁺18].

Be žymeklio sistemoje naudojamos kamerų sistemos skirstomos į dvi kategorijas: kuriančios gylio žemėlapi (angl. depth map) ir nekuriančios gylio žemėlapi. Neturint gylio informacijos dėl šešėlių, neidealios apšvietimo, atspindžių ir netvarkingo fono yra susiduriama su atpažinimo problemomis [CEC⁺18].

Kūno modelį sudaro žmogaus skeletas, aprašomas kaip rinkinys sąnarių, sujungtų kaulais tarpusavyje. Skeletas yra parametrizuojamas, [CEC⁺18] iš to kyla kompiuterinės regos uždavinys: kokį kūno modelį geriau naudoti efektyvesniam, tikslesniam identifikavimui, kad išvengti netikroviškų ar netaisyklingų pozų ir judesių. Renkantys paveikslėlio požymius, yra problematiška pasirinkti tokius, kad būtų įmanoma nustatyti žmogaus dydį ir vietą paveikslėlyje. Šiai užduočiai yra daugybė algoritmų, bet jų efektyvumą įtakoja šešėliai, apšvietimo pasikeitimai, atspindžiai, antraeiliai judesiai fone [CEC⁺18]. Vien silueto nustatymas nepateikia informacijos ar žmogus stovi atsisukęs į kamerą ar nuo jos. Šią problemą gali išspęsti 3D rekonstrukcija kai yra kelios kameros, tačiau ji reikalauja didesnių skaičiavimo išteklių [CEC⁺18]. Atsiradus giliesiems tinklams ir jų gebėjimui greitai ir patikimai nustatyti atskiras kūno dalis ženkliai pagerėjo galimybės nelaboratorinėmis sąlygomis identifikuoti siluetus, be to, tapo įmanoma nustatyti didelės žmonių grupės pozas vienoje nuotraukoje [CSW⁺17].

Pastaraisiais metais yra atlikta nemažai tyrinėjimų siekiant išspęsti aukščiau paminėtas problemas. Pavyzdžiui, [JNB19] įgyvendino priekinio plano segmentavimo metodą, padėsiančio tiksliau nustatyti žmogaus siluetą, kuris susideda iš odos aptikimo, remiantis odos spalva ir šviesimi. Jie pasiūlė naują žmogaus kūno modelį aprašyti 5 pagrindiniais ir 7 papildomais taškais ir gautus rezultatus patikrino, naudodant „KTH Football Multiview“ duomenų rinkinį [JNB19]. [YSG⁺19] panaudojo OpenPose biblioteką, padidinusią sistemos patikimumą bei lankstumą, identifikuojant pagrindinius taškus. Gilaus mokymo karkaso pagalba pavyko automatizuoti paveikslėlio požymių parinkimą. Buvo panaudotas laike paskirstytas konvoliucinio neuroninio tinklo sluoksnis su tikslu aptikti sąryšį tarp pagrindinių taškų viename kadre bei ilgos trumpalaikės atminties modelis (angl. long short-term memory (LSTM)) siekiant išsaugoti atmintyje atrastus sąryšius prieš tai buvusiuose kadruose. Taip pavyko sumažinti pagrindinių taškų atpažinimo klaidų skaičių, naudodant LSTM atmintyje išsaugotus kadrus vaizdo triukšmo mažinimui. Buvo padaryta išvada, kad pasiūlytas metodas jogos pozai nustatyti nereikalauja kameros, gebančios rinkti gylio informaciją, užtenka RGB kameros [YSG⁺19].

2018 metais buvo paskelbtas straipsnis su panašiu tikslu, padėti žmonėms taisyklingai atlikti pratimus [RSM⁺18]. Autoriai pateikė sprendimą, gebantį identifikuoti ir įvertinti žmogaus pozą nepriklausomai nuo žmogaus dydžio ir kūno sudėjimo. Jų sistema yra suprojektuota darbui su

keliais informacijos šaltiniais (kameromis), tačiau sprendžiant atpažinimo uždavinį buvo naudojama tik viena kamera ir supaprastintas kūno modelis (mažiau identifikavimo taškų). Yra pastebėta, kad sistema su viena kamera blogai identifikuoja sąnarius kai įvyksta zonų persidengimas (okliuzija), dėl ko gali suklysti pozos identifikavimo algoritmas [RSM⁺18].

Darbo tikslas – analizuojant veiksmų atpažinimo sistemas pereinamiesiems pratimams, pasiūlyti ir sukurti žmogaus pozos taškų pagrindu grįstus vaizdo įrašų klasifikavimo giliuosius neuroninius modelius ir jų realizacijas korektiškam veiksmų identifikavimui.

Darbo uždaviniai:

1. Atlikti vaizdo atpažinimo algoritmų, gebančių atpažinti judesius, mokslinės literatūros analizę.
2. Sukurti naują duomenų rinkinį pagal pratimus naudojamus kineziterapijose.
3. Atlikti tyrimą apmokant giliuosius neuroninius tinklus klaidingai atliekamų pratimų identifikavimui ir tyrimo metu pasiūlyti tinkamą algoritmą ir jo realizaciją.
4. Papildyti ir praplėsti giliuosius neuroninius modelius spręsti daugelio tikslų klasifikavimo uždavinius.

Uždavinio aktualumas

Potencialius sistemos, veikiančios kompiuterinės rega pagrindu, suinteresuotus vartotojus galima būtų suskirstyti į tokias grupes:

Potencialių vartotojų grupės:

1. Profesionalūs sportininkai
2. Pacientai su judesio sutrikimais
3. Žmonės besilaikantys sveikos gyvensenos principo

Potencialios taikymo aplinkos:

1. Sporto centrai
 - 1.1. Treniruoklių salė
 - 1.2. Grupinių užsiėmimų salės
2. Medicinos/reabilitacijos centrai
 - 2.1. Individualių/grupinių užsiėmimų patalpos/salės
3. Namų aplinka
 - 3.1. Gyvenamasis kambarys

Buvo identifikuotos kelios suinteresuotos grupės. Viena iš grupių yra Santaros klinikos gydytojai. Su keliais jų darbuotojais buvo atlikta konsultacija, tuo metu Santaros klinikos vyresniojo gydytoju rezidentu Laimonu Uščinu ir Santaros kineziterapeute, turinčia reabilitacijos magistro laipsnį, Silvija Reimeryte. Bendrai buvo įvardintos sritys, kuriose gydytojams reikėtų pagalbos arba gydytojai galėtų pasikliauti automatinėmis priemonėmis gauta informacija, pavyzdžiui:

1. Asmeninio lankstumo įvertinimas;
2. Jėgos pratimo atlikimo technikos įvertinimas;

3. Pusiausvyros treniruočių atlikimo technikos įvertinimas;
4. Funkcinis testavimas (pusiausvyros, koordinacijos);
5. Laikysenos įvertinimas (Pečių lanko ir stuburo linkių);

1 lentelė. Tinetti pratimai ir jų kodavimas

Unikalus numeris	Pavadinimas
Tinetti-1	Pusiausvyros laikymas sėdint
Tinetti-2	Atsistojimas iš sėdimos padėties
Tinetti-3	Bandytas atsistoti
Tinetti-4	Pusiausvyros laikymas pirmąsias 5 sek. po atsistojimo
Tinetti-5	Pusiausvyros laikymas stovint
Tinetti-6	Stumtelėjimas (paciento kojos suglaustos) 3 k.
Tinetti-7	Stumtelėjimas (pacientas užsimerkęs)
Tinetti-8	Apsisukimas 360° kampu
Tinetti-9	Atsisėdimas

Gydytojai paaiškino, kad raumenų jėgos testavimą galima įsivertinti naudojantis Lovett vertinimo metodu. Lovett metodą sudaro penkiabalė sistema:

0. – nėra raumens susitraukimo.
1. – matomas ar apčiuopiamas tik raumens susitraukimas.
2. – sugebama atlikti judesį pilna amplitude pašalinus gravitacinę jėgą.
3. – sugebama atlikti judesį pilna amplitude prieš gravitacijos jėgą.
4. – sugebama atlikti judesį pilna amplitude prieš gravitacijos jėgą su nedideliu pasipriešinimu.
5. – sugebama atlikti judesį pilna amplitude prieš gravitacijos jėgą su maksimaliu pasipriešinimu.

Šiam metodui taikyti, reikia fiksuoti kūno dalis, kur prasideda raumenys (proksimalinę testuojamos galūnės dalį) bei įsitikinti, ar judesys atliekamas per reikiamą sąnarį. Iš pateiktų lygių ir metodo taikymo taisyklių, galima suprasti, kad šis metodas negali būti pilna apimtimi panaudotas kompiuterinės regos sistemoje, nes iš vaizdinės medžiagos neįmanoma nustatyti kokia gravitacinė jėga yra naudojama testuojant žmogų bei nustatyti jo raumenų pradžios.

Pusiausvyra yra skirstoma į statinę (stovint) ir dinaminę (judant). Pusiausvyros ir eisenos vertinimui atlikti naudojamas Tinetti testas (žr. Priedas 1), BERG pusiausvyros vertinimo skalė [Arm16].

Lentelėje 1 yra pateikiami Tinetti pratimų pavadinimai ir atitinkamas pratimo kodas, kuris bus naudojamas toliau darbe, norint įvardinti Tinetti pratimą.

1. Gilieji neuroniniai tinklai

Šiame skyriuje yra apžvelgiami keli giliojo mokymo neuroniniai tinklai.

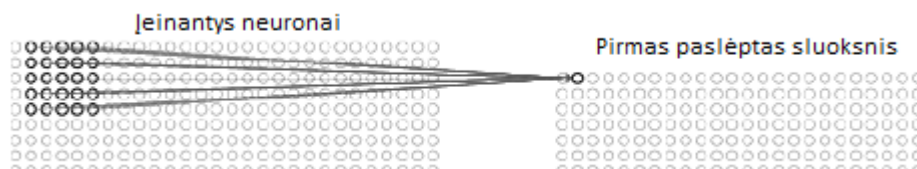
1.1. Konvoliuciniai neuroniniai tinklai

Dabartinių konvoliucinių neuroninių tinklų (angl. convolutional neural network) architektūros pradininkais galima laikyti Yann LeCun ir Léon Bottou, kurie dar 1998 metais pristatė neuroninių tinklų architektūrą pavadinimu LeNet 5, skirtą atpažinti ranka užrašytus žodžius bei skaitmenis [LBB⁺98]. Tuo metu jiems pavyko pasiekti užrašų atpažinimą 99,2% tikslumu, naudojant tikrinimui MNIST duomenų rinkinį.

Konvoliuciniai neuroniniai tinklai (CNN/ConvNet) yra vienas iš giliojo mokymo modelių, naudojamų apdoroti duomenis, kurių formatą galima išreikšti matricos pavidalu [YND⁺18]. Iš esmės tai yra tinklas sudarytas iš kelių tipų sluoksnių: konvoliucijos (angl. convolution), sutelkimo (angl. pooling) ir pilnai sujungto (angl. fully connected). Konvoliucijos ir sutelkimo sluoksniai atsakingi už požymių žemėlapių gavimą (angl. feature map extraction), o pilnai sujungtame sluoksnyje vyksta gautų požymių klasifikavimas [YND⁺18]. Konvoliuciniame neuroniniame tinkle turi būti bent vienas konvoliucijos sluoksnis, kuriame vyksta konvoliucijos operacija vietoje standartinės matricos daugybos operacijos [GBC16].

CNN tinklų privalumas prieš paprastus neuroninius tinklus su pilnai sujungtais sluoksniais yra tame, kad jie atsižvelgia į paveikslėlių struktūrą (vaizdo taškai (angl. pixels), kurie yra šalia vienas kito, ir vaizdo taškai, kurie yra nutolę vienas nuo kito, traktuojami skirtingai). Taip yra paspartinamas tinklo apmokymas, todėl galima naudoti daugiau paslėptų sluoksnių (angl. hidden layers) ir atrasti sudėtingesnes priklausomybes [Nie15].

Konvoliuciniai neuroniniai tinklai yra paremti trimis pagrindinėmis idėjomis: vietiniais priimamaisiais laukais (angl. local receptive fields), bendrais svoriais (angl. shared weights) ir sutelkimu (angl. pooling) [Nie15]. Vietiniais priimamaisiais laukais yra vadina įeinamojo sluoksnio vaizdo taškų grupė, kuri yra sujungiamas su paslėpto sluoksnio neuronais. Kiekvienas vietinis priimamasis laukas yra sujungiamas su atskiru paslėpto sluoksnio neuronu, kaip pademonstruota paveikslėlyje 1 [Nie15].



1 pav. Lokalus matymo lauko sujungimas su paslėpto sluoksnio neuronu

Taip yra „pereinama“ (vykdoma konvoliucijos operacija) per visą įeinamojo sluoksnio vaizdo taškų masyvą (taškų matricą) ir formuojamas duomenų masyvas tolimesniam sluoksniui. Yra galimybė nustatyti kokio dydžio vietinių priimamųjų laukų dydį sujungti su sekančio sluoksnio neuronu [Nie15]. Tai nustatoma filtro, dar kitaip vadinamo filtro branduolio (angl. kernel), parametru pagalba. Pats filtras yra apibrėžiamas kaip parametru (angl. weights) masyvas. Reikia

pažymėti, kad filtro ir įeinančių duomenų gylis (matricos sluoksnių skaičius) turi būti vienodas, kad galima būtų įvykdyti matematinės matricių operacijas. „Perėjimą“ galima konfigūruoti, nustatant skirtingo dydžio žingsnį (angl. stride). Tai reiškia, kad jeigu žingsnis lygus 1, tada filtras pasislenka per vieną vaizdo tašką ir yra įvykdoma konvoliucijos operacija, jeigu lygus 2, tada per du taškus ir taip toliau. Po kiekvieno konvoliucinio sluoksnio suformuotas aktyvacijų žemėlapis visada yra mažesnis už įeinančių duomenų matricą ir pirminiuose tinklo sluoksniuose jis mažėja greičiau negu yra norima. Tam, kad išvengti to, gauta matrica po konvoliucijos operacijos yra praplečiama „aplink“ apskaičiuotus duomenis pridėdant papildomus stulpelius ir eiles, bei jas užpildant nuliais. Kiek reikia praplėsti gautą matricą nurodo papildymo (angl. padding) parametras. Norint apskaičiuoti išvesties duomenų dydį po konvoliucinio sluoksnio galima pasinaudoti Formule 1.

$$O = ((W - F + 2P)) / (S + 1) \quad (1)$$

W – įvesties duomenų dydis,

F – filtro branduolio reikšmė,

P – papildymo reikšmė,

S – žingsnio reikšmė,

O – išvesties duomenų dydis.

Įvardinti konfigūruojami dydžiai (filtras, žingsnis, papildymas) sudaro neuroninio tinklo hiperparametrus (angl. hyperparameters), kurios reikia nurodyti prieš paleidžiant tinklo apmokymą. Taip pat, yra galimybė nurodyti filtrų skaičių (angl. filter count), kuriuos reikia naudoti. Kuo didesnis laipsnis, tuo yra galingesnis modelis, tačiau kartu didėja ir tinklo persimokymo rizika. Dažniausiai yra pradama nuo mažesnio kiekio filtrų pradiniuose tinklo sluoksniuose bei progresyviai didinamas filtrų skaičius vėlesniuose sluoksniuose.

Konvoliucija. Tai yra tam tikra tiesinė operacija, kurios tikslas yra gauti požymių žemėlapi (angl. feature map), dar vadinamu aktyvacijų žemėlapiu (angl. activation map) [YND⁺18]. Matematinė išraiška atrodo kaip nurodyta Formulė 2.

Naudojant vieną funkciją (vienmatėje sistemoje):

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \quad (2)$$

Naudojant dvi funkcijas (dvimatėje erdvėje):

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(m,n)K(i - m, j - n) \quad (3)$$

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i - m, j - n)K(m,n) \quad (4)$$

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i + m, j + n)K(m,n) \quad (5)$$

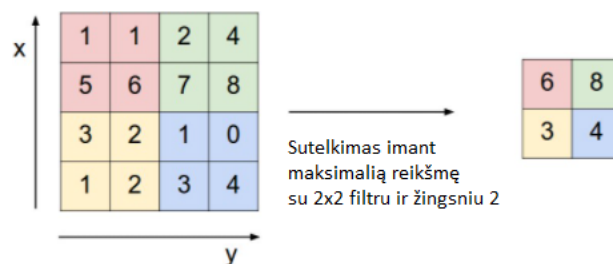
Konvoliucijos sluoksnis

Konvoliucinis sluoksnis yra pagrindinis sluoksnis atskiriantis šio tipo tinklus nuo kitų. Jame yra atliekama konvoliucijos operacija tarp įvesties duomenų ir filtro.

Sutelkimo sluoksnis

Sutelkimo (angl. pooling) sluoksniai dažniausiai yra naudojami po konvoliucinių sluoksnių. Jų tikslas yra apibendrinti informaciją po konvoliucinių sluoksnių ir gauti suglaustą požymių žemėlapi (angl. condensed feature map) apskaičiuota pagal tam tikrą metodą [Nie15]. Taip yra pagreitinamas tinklo apmokymas ir apsisaugoma nuo tinklo persimokymo (angl. overfitting).

Sutelkimas imant maksimalią reikšmę (angl. max pooling). Tai yra metodas pagal kurį yra vykdomas duomenų supaprastinimas (pav. 2). Tikslas yra surasti požymį ir jo apytikslę santykinę poziciją kitų požymių atžvilgiu. Tai yra pasiekama po kiekvieno požymio žemėlapio gavimo, sutelkimo sluoksnyje surandant didžiausią reikšmę gautoje matricoje ir ją užrašant į suglaustą požymių žemėlapi. Tuo siekiama sumažinti parametrų skaičių sekančiuose sluoksniuose neprarandant svarbios informacijos [Nie15]. Taip pat yra naudojamas „L2 norm pooling“ metodas, kuris vietoj didžiausios reikšmės radimo, ištraukia kvadratinį šaknį iš apskaičiuotos kvadratinės sumos gautoje matricoje [Nie15], bei sutelkimo vidurkinant (angl. average pooling) metodo atveju yra randamas vidurkis kiekvienoje gautoje matricoje ir reikšmė užrašoma į suglaustą požymių žemėlapi.



2 pav. Sutelkimo sluoksnio, imant maksimalią reikšmę, pavyzdys. Yra pavaizduota 4x4 dydžio matrica, 2x2 dydžio filtras ir žingsnis su reikšmė 2. Kiekviename 2x2 dydžio filtre yra randama didžiausia reikšmė ir ji įrašoma į išvesties matricą ir taip suformuojama 2x2 dydžio matrica su didžiausiomis reikšmėmis iš kiekvieno filtro

Aktyvacijos funkcijos

Vienas iš neuroninio tinklo kūrimo etapų yra neuronų aktyvacijos funkcijos pasirinkimas. Paprasčiausia aktyvacijos funkcija vadinama tiesine aktyvacija kai transformacija visiškai netaikoma. Tinklą, kuriame naudojamos tik tiesinės aktyvacijos funkcijos, yra lengva apmokyti, tačiau tokie tinklai negali išmokti sudėtingų požymių žemėlapinių. Tiesinės aktyvacijos funkcijos vis dar naudojamos išvesties sluoksnyje tinkluose, kurie sprendžia regresines problemas.

Tačiau pirmenybė teikiama netiesinėms aktyvacijos funkcijoms, kadangi jos leidžia neuro-
nams išmokti sudėtingesnes duomenų struktūras, pavyzdžiui, sigmoido, hiperbolinė tangento,

ReLU [GBC16], Leaky ReLU [LWW⁺19], Parametric ReLU [HZR⁺15], ELU [CUH15], Swish [RZL17].

Svarbu paminėti, kad naudojant tam tikras aktyvacijos funkcijas su daug vidinių sluoksnių neuroniniuose tinkluose gali pasireikšti nykstančių gradientų (angl. vanishing gradient) problemos arba sprogtančių gradientų (angl. exploding gradient) problemos. Nykstančių gradientų problema atsiranda tada kai vykdomas atgalinio sklidimo (angl. backpropagation) algoritmas ir skaičiuojamos išvestinės, jų reikšmė būna labai maža (arti nulio) ir dėl to tinklo pradinėse sluoksniuose svorių atnaujinimas vyksta neefektyviai arba iš viso neįvyksta. Sprogstančių gradientų problema reiškia, kad įvyksta staigūs/dideli svorių atnaujinimai kas padaro tinklą nestabilų ir jis nesugeba apsimokyti naudojant esamus duomenis [Nie15]. Šių problemų sprendimas yra eksperimentiškai parinkti labiau tinkančias aktyvacijos funkcijas.

Sigmoido aktyvacijos funkcija

Funkcijos įvesties reikšmės yra transformuojamos į reikšmes iš intervalo [0;1].

$$g(x) = \sigma(x) = (1 + e^{-x})^{-1} \quad (6)$$

Hiperbolinio tangento (sutrumpinant - tanh) aktyvacijos funkcija

Funkcijos įvesties reikšmės yra transformuojamos į reikšmes iš intervalo [-1;1], kur įvesties reikšmės didesnės už 1 yra užrašomas kaip 1 ir mažesnės už -1 yra užrašomos kaip -1.

$$g(x) = \tanh(x) \quad (7)$$

Sigmoido ir tanh aktyvacijos funkcijos yra glaudžiai susijusios. Jų sąryšį galima pamatyti iš formulės 8.

$$\tanh(x) = 2\sigma(2x) - 1 \quad (8)$$

Sigmoido ir tanh funkcijos pasižymi sodrumo (angl. saturate) problema. Kai funkcijos parametras yra labai mažas, arba labai didelis, funkcija tampa labai plokščia ir nejautri mažiems pokyčiams. Todėl esant dideliame sodrumui, gradientu pagrįstas apmokymas [Cui18] gali būti apsunkintas. [GBC16]

Dalimis tiesinė funkcija

Dalimis tiesinė funkcija (angl. Rectified linear activation unit), dar vadinama tiesiog „ReLU“, yra netiesinė aktyvacijos funkcija, kuri elgiasi kaip tiesinė kai funkcijos argumentai yra teigiamos reikšmės, o kai argumentai yra neigiami, funkcijos reikšmė visada yra nulis. Jos dėka galima apmokinti tinklą atpažinti sudėtingas priklausomybes duomenyse. [GBC16]

$$g(x) = \max(0, x) \quad (9)$$

„Mirusio ReLU“ problema (angl. dead ReLU problem). Esant neigiamiems ar nuliams sluoksnio įvesties duomenims, ReLU grąžinama reikšmė visada yra nulis. Problema atsiranda norint ištaisyti apmokymo klaidą atgalinio sklidimo (angl. backpropagation) būdu, kuriuo metu skaičiuojama funkcijos išvestinė ir bandoma atnaujinti svorius (angl. weights). Kadangi funkcijos išvestinė būna nulis, algoritmas negali atnaujinti svorių ir persimokyti, todėl sakoma, kad neuronas „mirė“. [LSS⁺19]. Šiai problemai išspręsti sukurta „Leaky ReLU“ aktyvacijos funkcija.

„Leaky/Parametric ReLU“ aktyvacijos funkcija

Leaky ReLU (LReLU) yra modifikuota ReLU funkcijos versija. Jos esmė yra, kad įvesties duomenų reikšmė padauginama iš mažos teigiamos reikšmės $\epsilon < 1$. Dėl to aktyvacijos funkcijos rezultatas gautas su neigiamomis įvesties duomenų reikšmėmis nėra konvertuojamas į nulį ir vykdant atgalinį sklidimą įmanoma apskaičiuoti funkcijos išvestinę ir permokinti tinklą. [LWW⁺19]

$$g(x) = \max(\epsilon \cdot x, x), \quad \text{kur } \epsilon < 1 \quad (10)$$

Parametric ReLU (PReLU) esmė ta pati kaip ir Leaky ReLU, tik su vienu skirtumu, ϵ yra išmokstamas kintamasis, kuriuo reikšmė yra randama tinklo mokymosi metu. [HZR⁺15]

EkspONENTINIS TIESINIS VIENETAS

EkspONENTINIS TIESINIS VIENETAS (angl. Exponential linear unit (ELU)) yra aktyvacijos funkcija, kuri turi papildomą globalią konstantą α , kurios reikšmė turi būti teigiama. Formulė 11 atvaizduoja ELU funkciją.

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \quad \text{kai } \alpha > 0 \quad (11)$$

Skirtingai nei ReLU, ELU konverguoja į nulį greičiau ir geba pateikti tikslesnį rezultatą. Dėl to ELU naudojimas leidžia pagreitinti tinklo apmokymą, pasiekti geresnį apibendrinimą (angl. generalization) apmokant neuroninį tinklą su daugiau nei 5 sluoksniais [CUH15].

„Swish“

„Swish“ yra nemonotoniška, glotni funkcija, kurios reikšmių sritis yra apribota iš apačios ir tęsiasi iki begalybės. [RZL17] „Swish“ funkcija (formulė 12) turi parametą β , kuris yra nustatomas tinklo mokymosi metu.

$$f(x) = x \cdot \sigma(\beta \cdot x), \text{ kur } \beta - \text{išmokstamas parametras} \quad (12)$$

Yra funkcijos versijų, formulė 13, kuriuose parametras β yra nenaudojamas. Tokia funkcija vadinama „Swish-1“ ir atitinka SiLU (angl. sigmoid-weighted linear unit).

$$f(x) = x \cdot \sigma(x), \text{ kur } \sigma(x) - \text{sigmoido funkcija} \quad (13)$$

Tais atvejais kai parametras β yra nustatomas lygus vienetui, funkcija tampa tiesinė (formulė 14) ir atrodo kaip ReLU funkcija su teigiamais argumentais. [RZL17]

$$f(x) = x \quad (14)$$

„GELU“

Gauso klaidos tiesinio vieneto (angl. Gaussian error linear unit) aktyvacijos funkcija (formulė 15). Hendrycks ir Gimpel teigia, kad jų tikslas buvo apjungti ReLU, ELU, PReLU, išmetimo operacijų bei rekurentiniams neuroniniams tinklams atrasto „zoneout“ [KMK⁺16] metodo teigiamas savybes į vieną funkciją. Jie atsitiktinai dauginą įeinančius duomenis iš 0 arba 1 ir deterministiškai gauna aktyvacijos funkcijų rezultatus [HG16].

$$GELU(x) = x \cdot P(X \leq x) = x \cdot \Phi(x) = x \cdot \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right], \text{ kai } X \sim \mathcal{N}(0,1) \quad (15)$$

Kas apytiksliai lygu

$$f(x) = \frac{1}{2}x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}}(x + 0.044715x^3) \right] \right) \quad (16)$$

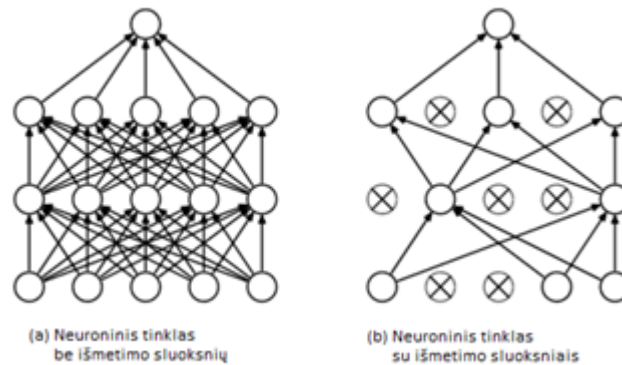
Arba

$$x \cdot \sigma(1.702x) \quad (17)$$

Atsitiktinio praretinimo transformacija

Tinklo apmokyme egzistuoja problema kai tinklas išmoksta per daug tiksliai nustatyti duomenų požymius. Ši problema vadinama persimokymo (angl. overfitting). Problema pasireiškia tada kai tinklas tampa pernelyg sudėtingas sprendžiamai problemai, kitaip sakant, pernelyg gerai išmoksta atpažinti duomenų požymius iš mokymui skirtu duomenų rinkinio. Jeigu tinklui duoti bet kokius naujus duomenis, jis nesugebės priiminti sprendimų su tokiu pačiu tikslumu, kurį rodė su apmokymo duomenimis. Tokių situacijų išvengimui sugalvotas atsitiktinio praretinimo

(angl. dropout) transformacija, kurios paskirtis yra apmokymo metu nenaudoti atsitiktinį skaičių aktyvacijos duomenų. Paveikslėlyje 3 galima pamatyti kaip supaprastėja neuroninio tinklo ryšių struktūra, pritaikius atsitiktinio praretinimo transformaciją. [WG15]



3 pav. Atsitiktinio praretinimo transformacijos pavyzdys. Paveikslėlis (a) parodo neuroninį tinklą prieš transformaciją, kiekvienas neuronas yra sujungtas su kitais neuronais. Paveikslėlyje (b) yra pavaizduotas tinklas po pritaikytos transformacijos [SHK⁺14]

Nuostolių funkcijos

Nuostolių funkcijų paskirtis yra nustatyti mokymosi tikslą bei pamatuoti kaip gerai modelio prognozės atitinka įvesties duomenų tikrąsias reikšmes. Jas galima būtų sugrupuoti į dvi kategorijas: 1) klasifikavimo ir 2) regresinės nuostolių funkcijos. Klasifikavimo problemoms spręsti įprastai yra naudojama kryžminės entropijos (angl. cross-entropy) nuostolių funkcija, formulė 18.

$$L(y, \hat{y}) = - \sum_{i=1}^n y_i \ln \hat{y}_i \quad (18)$$

Pilnai sujungtų sąryšių sluoksnis

Pilnai sujungtų sąryšių sluoksnis (angl. fully connected layer) sujungia kiekvieną prieš tai buvusio sluoksnio neuroną su kiekvienu neuronu sekančiame sluoksnyje. Jo tikslas yra gautus aukšto lygio požymių žemėlapius iš konvoliucinio ar apjungimo sluoksnių susieti su įvairiomis klasėmis, kurios yra pateiktos apmokymo duomenų rinkinyje. Prieš tai konvoliucinio ar apjungimo sluoksnio išvesties reikšmės yra ištiesinamos (angl. flatten) ir gaunamas vektorius. Tada prieš pat išvesties (angl. output) sluoksnį yra naudojama „Softmax“ aktyvacijos funkcija, kuri priima reikšmių vektorių $x \in \mathbb{R}^n$ ir grąžina vektorių su tikimybėmis $p \in \mathbb{R}^n$. [QKG19]

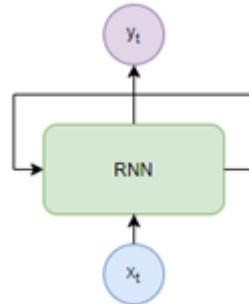
$$p = (p_1 \mid \dots \mid p_n), \text{ kur } p_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (19)$$

Pilnai sujungto sluoksnio rezultatas yra vektorius su tikimybėmis, kurių reikšmės pasiskirsto tarp 0 ir 1, o visų vektoriaus tikimybių suma yra lygi vienetui. Vektoriaus didžiausia reikšmė parodo, kuriai klasei, anot tinklo, priskiriamas tinklo įvesties duomuo.

1.2. Rekurentiniai neuroniniai tinklai

Rekurentiniai neuroniniai tinklai (angl. Recurrent neural networks (RNN)) yra dirbtinių neuroninių tinklų klasė, kurios idėją aprašė David Rumelhart 1986 metais [RHW86]. Rekurentinių neuroninių tinklų išskirtinis bruožas yra neuronų sujungimas, kuris suformuoja orientuotą grafą laiko sekoje.

Skirtingai nei paprastuose be grįžtamojo ryšio neuroniniuose tinkluose, kur duomenys yra perduodami į vieną pusę ir galimybės išlaikyti informacijos apie praities įvykį ar įvykių seką nėra, rekurentiniai tinklai turi savyje ciklišumą. Tai leidžia apdoroti tam tikro laikotarpio informaciją.



4 pav. Rekurentinio neuroninio tinklo koncepcija

Paveikslėlyje 4 pavaizduotas rekurentinio neuroninio tinklo koncepcija, kur t – laiko žingsnis. Tinklui yra paduodami duomenys už tam tikrą laikotarpį x_t , kuriems tinklas atlieka matematinės operacijas ir gauna tinklo išvesties reikšmę y_t . Papildomai yra apskaičiuojamas tinklo vidinės būsenos reikšmė h_t duotu laiku t , kuri paskui yra perduodama tinklui sekančiam laiko žingsniui $t + 1$. Matematinė išraiška tai užsirašo kaip nurodyta formulėje 20. [GBC16]

$$h_t = f_W(h_{t-1}, x_t) \quad (20)$$

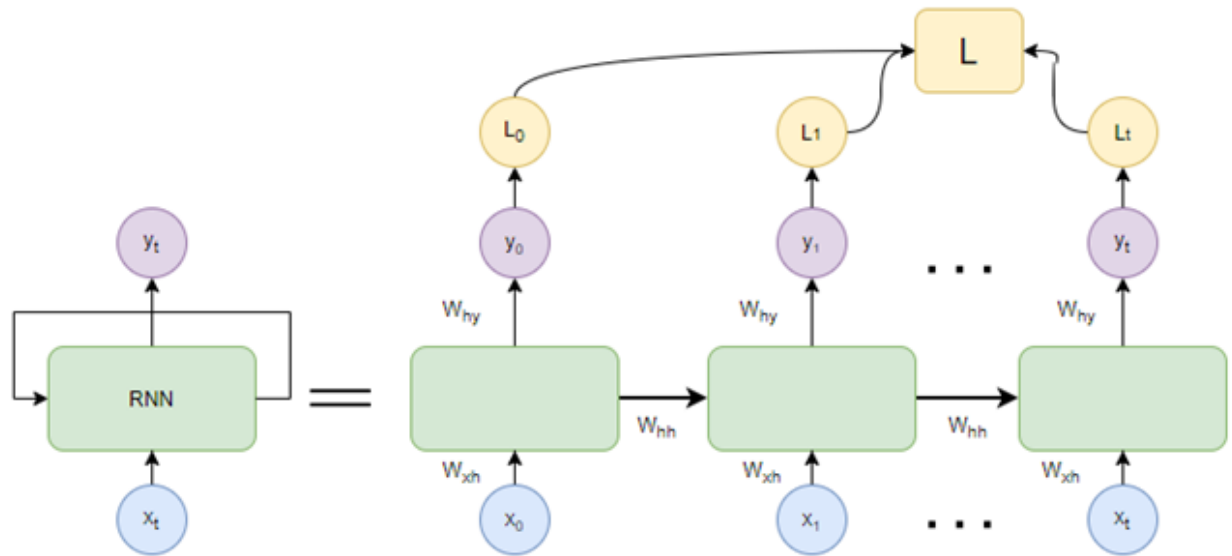
h_t – vidinė būseną,

f_W – funkcija, parametrizuota W parametrais,

h_{t-1} – sena vidinė būseną,

x_t – įvesties duomenų vektorius laiko žingsniu t .

Nurodyta funkcija f_W ir parametrų rinkinys, naudojami RNN tinkle, yra tos pačios ir nekinta skaičiuojant išvesties reikšmes visuose laiko žingsniuose.



5 pav. Paprasto rekurentinio neuroninio išskleisto tinklo koncepcija

Paveikslėlyje 5 yra pademonstruotas išskleistas RNN tinklas kiekvienu laiko žingsniu. Ji galima suprasti kaip t -kopijų to paties tinklą, kur iš vienos kopijos į kitą perduodama žinutė suformuota vidinės būsenos h_t pagrindu. W_{xh} yra pažymėti svoriai, kurie naudojami apskaičiuojant vidinę būseną kai paduodamos įeinančios reikšmės x_t . W_{hh} yra svoriai, naudojami apskaičiuojant dabartinę vidinę būseną remiantis prieš tai buvusia vidine būsena $h_t - 1$. W_{hy} yra svoriai naudojami apskaičiuoti rezultatą remiantis vidine būsena h_t . Naudojant kiekvieną rezultatą y_t galima apskaičiuoti nuostolį L_t kiekviename laiko žingsnyje, kuris vykdomas atgalinio sklidimo laiko žingsnyje (angl. backpropagation through time (BPTT)) [BMK⁺17] metu. Susumavus klaidas, gausime bendra nuostolį kokį turėjome vykdydami tinklo apmokymą. [GBC16]

Kiekviename atskirame laiko žingsnyje yra naudojamos tokios pačios svorių matricos, kitaip sakant, W t_0 žingsnyje yra lygi W t_1 žingsnyje.

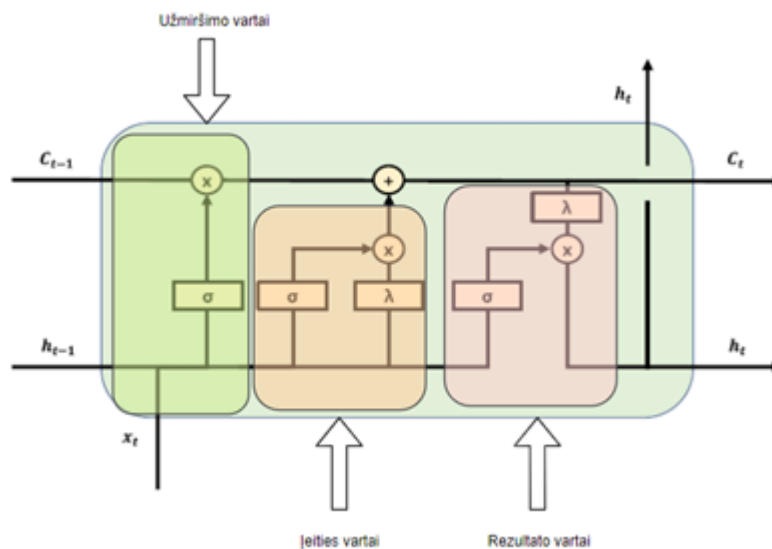
Rekurentinių neuroninių tinklų architektūros

Šiame poskyryje aprašomi ilgos trumpalaikės atminties modelio (angl. long short-term memory (LSTM)) bei sklendžių rekurentinių vienetų modelio (angl. Gated recurrent units (GRU)) RNN tinklo variacijos.

Ilgos trumpalaikės atminties modelis

Šios ilgos trumpalaikės atminties modelio (angl. long short-term memory (LSTM)) RNN tinklo variacija pirmą kartą buvo pristatyta Hochreiter ir Schmidhuber dar 1997 metais [HS97]. Šios architektūros atsiradimas buvo atsakas į nykstančių gradientų (angl. vanishing gradient) problemą bei trumpalaikės atminties problemą kai standartiniam RNN tinklui neišeidavo su pakankamu tikslumu susieti faktus, kurie yra toli vienas nuo kito nutolę laike [BMK⁺17]. Pastarajai problemai spręsti, LSTM architektūroje pasikeitė vidinio modelio struktūra ir atsirado 3 sklendės (angl. gates): atidarymo sklendė (angl. input gates), išvesties sklendė (angl. output gates) ir užmiršimo

sklendė (angl. forget gates) bei papildomi neuroninio tinklo sluoksniai. Paveikslėlyje 6 yra pavaizduota LSTM ląstelės struktūra. Linija pažymėta C_{t-1} ir C_t yra ląstelės būsenos. Šia linija vyksta informacijos srautas, kuris gali būti pakoreguotas loginio vieneto pagalba. Atidarymo sklendėje yra nustatoma kokia nauja informacija bus saugoma ląstelės būsenoje. Tai įvyksta dviem etapais. Pirmiausia sigmoido sluoksnyje nustatoma kurias reikšmes reikia atnaujinti, tangento sluoksnyje nustatoma į kokias reikšmes potencialiai galima pakeisti reikšmes ir tada tų dviejų sluoksnių gauti rezultatai yra sudauginami ir toliau panaudojami apskaičiuojant C_t reikšmę. Užmiršimo sluoksnyje yra nustatoma kurią informaciją galima „užmiršti“. Už tai atsakingas atskiras sigmoido sluoksnis, grąžinantis reikšmes nuo 0 iki 1 kiekvienai reikšmei iš C_{t-1} būsenos, kur 0 reiškia „visiškai pamiršti“, ir 1 – išlaikyti reikšmę. Gautos reikšmės taip pat yra panaudojamos C_t apskaičiuoti. Rezultato sluoksnyje yra taikomi dar vienas sigmoido ir tangento sluoksniai, kurie yra naudojami vidinės būsenos apskaičiavimui. Gautos reikšmės perduodamos sekančiam laiko žingsniui. [HS97]



6 pav. LSTM ląstelės struktūra

Skendžių rekurentinių vienetų modelis

Skendžių rekurentinių vienetų modelio (angl. Gated recurrent units (GRU)) RNN variacija veikia panašiai kaip ir LSTM, kadangi buvo kurta su tikslu spręsti trumpalaikės atminties problemą RNN modeliuose. Vietoj to, kad informacijos reguliavimui naudoti ląstelės būseną, GRU yra naudojamos paslėptos būsenos. Ir nuo 3 skendžių GRU liko tik 2 sklendės: grįžimo į pradinę būseną arba atstatymo (angl. reset) sklendė ir atnaujinimo (angl. update) sklendė. Atnaujinimo sklendė elgiasi panašiai kaip užmiršimo ir atidarymo LSTM sklendės. Joje yra nusprendžiama kokią informaciją išmesti ir kokią naują pridėti. Atstatymo sklendėje yra nustatoma kiek praeties informacijos reikia užmiršti [CVG⁺14].

GRU pasižymi spartesne greitimeika bei mažesniu atminties sunaudojimu kadangi turi mažiau parametrų dėl pasikeitusios vidinės struktūros. Tuo pačiu metu GRU rodo pakankamai gerus apmokymo rezultatus panaudojant didelius duomenų rinkinius [LDK⁺20].

2. Veiksmo atpažinimo sistemos

Veiksmo atpažinimo užduotys buvo pradėtos tirti nuo pat kompiuterinės regos atsiradimo ir iki šiol yra aktyviai tiriamos, kadangi šia technologija paremtos sistemos turi galybę įvairių pritaikymų (vaizdo indeksavimas, biometrijos panaudojimas, medicinos, stebėjimo bei apsaugos sistemose [RAM16]). Jų tikslą galima būtų suvesti į žmogaus veiksmo atpažinimą iš sensorių arba vaizdo medžiagos. Žmogaus veiksmus galima suskirstyti į keturias grupes [AR11]:

1. Gestai. Tai yra paprasčiausi žmogaus kūno judesiai, tokie kaip „rankos ištiesimas“ ar „kojos pakėlimas“.
2. Veiksmai. Yra vieno žmogaus gestai, kurie tęsiasi tam tikrą laiką, pavyzdžiui „ėjimas“.
3. Sąveikos. Apima dviejų ir daugiau žmonių sąveikos tarpusavyje arba/ir su kitais objektais. Vienas iš pavyzdžių „važiuojant dviračiu partrenkti kitą žmogų“.
4. Grupiniai. Kaip ir sąveikų atveju, apima du ir daugiau žmonių arba/ir objektų, tačiau atliekančius tą patį veiksmą, pavyzdžiui, „grupės plaukimas baseino takelyje“.

Veiksmo atpažinimo sistemos (VAS) skirstomos į paremtas sensoriais, paremtas ne vaizdo sensoriais bei multi-modalines sistemas [ACN⁺20].

2020 metais buvo atlikti 2 mokslinės literatūros apžvalgos tyrimai [ACN⁺20; JKA⁺20], kur buvo apžvelgtos pagrindinės problemos su kuriomis susiduriama, naudojant kompiuterinę regą bei naudojami veiksmo atpažinimo algoritmai. Šiame skyriuje yra aprašomi veiksmo atpažinimo problemos ir metodai, kurie galėtų būti taikomi iškeltiems uždaviniams įgyvendinti, papildant mokslo straipsniuose paskelbtais naujais metodais.

2.1. Veiksmo atpažinimo sistemų problemos

Apmokymo ir testavimo duomenų pasirinkimas

Tai yra svarbi sistemos sudedamoji dalis, kadangi nuo jos gali stipriai priklausyti modelio tikslumas. Dažniausiai kaip įvesties duomenys veiksmo atpažinimo uždaviniuose yra naudojamos RGB, RGB su gylio informacija ar žmogaus pozos (angl. skeleton) taškų informacija.

- RGB – pateikia labai daug vaizdinės informacijos. Iš kitos pusės, klasifikatoriaus sprendimą gali įtakoti visai kitokie požymiai negu buvo tikimasi, pavyzdžiui apšvietimas ar galinio plano objektai scenoje.
- RGB su gylio informacija – privalumas, kad nepriklauso nuo besikeičiančio apšvietimo, galima lengviau atpažinti galinio plano objektus scenoje, bet suteikia 3D informaciją apie filmuojamą sceną. Tačiau, problemos gali atsirasti problemos gylio žemėlapiuose, kurių priežastis gali būti triukšmas. Tokiu atveju reikia taikyti papildomus metodus, kad pašalinti arba sumažinti triukšmą.
- Žmogaus pozos (angl. skeleton) taškų informacija – gali būti gauta skirtingais būdais iš paprastų RGB duomenų arba ir kartu panaudojus gylio informaciją. Tačiau šios informacijos kokybė labai priklauso nuo sąlygų, kuriuose buvo filmuojamas pradinis duomenų šaltinis,

pavyzdžiui, ar nebuvo vaizdo triukšmo aplinkoje ar okliuzijos elementų ar kitų problemų [ACN⁺20].

Požiūrio taškų variacijos

Yra žinoma požiūrio taškų variacijos problema (angl. variation in viewpoint), kad dauguma metodų daro prielaidą, jog veiksmas atliekamas iš fiksuoto taško. Tas pats galioja ir judesio šablonams, kas apsunkina tokių veiksmų ir judesių nustatymą. Šioms problemoms spręsti buvo pasiūlyta naudoti kelių kamerų informaciją, modeliuoti 3D kūno pozą, naudoti nuo požiūrio nepriklausomus požymius. Tačiau buvo pastebėta, kad norint nustatyti pasiūlytų metodų efektyvumą reikia naudoti informaciją gautą tikroviškoje aplinkoje, o ne gautą specialiai ar dirbtinai pritaikytose sąlygose [ACN⁺20].

Okliuzijos problema

Kompiuterinės regos kontekste okliuzija (angl. occlusion) suprantama kaip laikinas kūno dalių uždengimas kitais objektais (daiktais arba žmonėmis). Egzistuojančias okliuzijos problemas galima suskirstyti į 3 kategorijas [JKA⁺20]:

1. Paties savęs uždengimas (angl. self-occlusion). Tokia situacija gali pasitaikyti kai iš vieno taško perspektyvos dalis kūno dalių yra uždengiama kitų kūno dalių. Pavyzdžiui, žiūrint iš tam tikro taško ir atliekant Tinetti-6 užduotį, žmogaus antra koja/ranka nesimatyti.
2. Susibūrimo problema (angl. crowd issue). Pavyzdžiui, vykstant grupinei treniruotei, jeigu kamera iš priekio, tada pirmoje eilėje stovinčių žmonių kūnai galėtų uždengti žmonių, stovinčių antroje eilėje, kojas ar kitas kūno dalis ir t.t.
3. Uždengimas sąlygotas kito objekto. Tokia situacija gali pasitaikyti, atliekant Tinetti-1, Tinetti-2, Tinetti-3 ir panašiose užduotyse, kur bus naudojama kėdė ir jeigu kamera bus pastatyta žmogui už nugaros. Tada kėdė gali uždengti dalį kūno dalių, tokių kaip kojos, dalį stuburo ir t.t.

Netvarkingo fono problema

Netvarkingo fono (angl. cluttered background) problema atsiranda dėl galinės scenos triukšmo. Gaunant globalius požymius iš nekontroliuojamos aplinkos vaizdo įrašų, per klaidą gali būti paimtas galinės scenos triukšmas kaip perteklinė informacija. Tokios informacijos įtraukimas sąlygoja prastesnę tinklo apmokymo spartą. Norint išvengti panašių problemų, požymių gavimui galima naudoti metodus, kurie veikia taškų identifikavimo principu. Taip pat lokalių požymių nustatymo technikos, tokios kaip (angl. space time interest points) ir krašto žemėlapiai (angl. edge maps), pasirodė efektyviai galinčios spręsti netvarkingo fono problemas atskirais atvejais. Nepaisant to, jie yra nepakankami, norint išspręsti globalią problemą. Papildomai galima pasitelkti IR kameras triukšmų pašalinimui, tačiau tais atvejais kai kamera filmuoja tik RGB formatu buvo pasiūlyti segmentavimo metodai, veikiantys spalvos pagrindu (angl. color-based) ir regiono pagrindu (angl. region-based) [JKA⁺20].

Tarp klasių panašumo ir kintamumo klasės viduje problemos

Yra žinomos tarp klasių panašumų ir kintamumo klasės viduje (angl. inter-class similarity and intra-class variability) problemos. Kintamumo klasės viduje problemą susidaro dėl to nes tam pačiam žmogui yra labai sudėtinga atlikti identišškai tą patį veiksmą. Jeigu tą patį veiksmą paprašoma atlikti skirtingų žmonių, galima pastebėti, kad visi jie atliks ją skirtingai dėl vienokių ar kitokių priežasčių. Yra veiksmų, kurie pagal požymius labai panašūs į kitus, tačiau iš esmės yra skirtingi, tai yra vadinama tarp klasių panašumo problema. Šioms problemoms spręsti buvo pasiūlyti tikimybiniai, tokie kaip paslėpti Markovo modeliai, dinaminio Bayesian (angl. dynamic Bayesian) metodas, baigtiniai automatai (angl. finite state machines), sąlyginiai atsitiktiniai laukai (angl. Conditional Random Fields). Jie remiasi vertinimais laiko pagrindu. Taip pat, dinaminio laiko iškraipymo (angl. dynamic time warp) modeliai, jų esmė yra palyginti dvi laiko sekas, kuriose veiksmai gali būti atliekami skirtingu greičiu. Pavyzdžiui, dviejų skirtingais greičiais einančių žmonių palyginimas, net jeigu buvo pagreitinimai arba sulėtėjimai pas vieną iš žmonių. Šitas metodas nereikalauja daug duomenų apmokymui, tačiau gali pareikalauti didelių skaičiavimo resursų, jeigu susiduriama su didele kintamumo vidinėje klasėje problema [ACN⁺20].

Priede 1 pateikti Tinetti testo kai kurie pratimai yra panašūs vieni į kitus, todėl tikėtina, kad šios problemos gali pasireikšti.

2.2. Požymių nustatymo būdai

Sprendžiant veiksmų atpažinimo uždavinį galima įvairiais būdais nustatyti skirtingo lygio požymius (angl. features). Vieni mokslininkai buvo pasiūlę naudoti holistinius metodus. Jie gali būti suskirstyti į du pogrupius: 1) atpažinimas, remiantis formos informacija, pavyzdžiui, formos užmaskavimu (angl. shape mask) ir žmogaus siluetu, ir 2) atpažinimas, remiantis formos ir visuotine judesio informacija. Vienu ar kitu atveju nustačius reikalingus pikselius, jie visi yra paimami ir panaudojami deskriptoriaus skaičiavimui. Kitaip sakant, šie metodai tinka tais atvejais kai nėra poreikio identifikuoti atskiras žmogaus kūno dalis. Pirmos grupės metodai susiduria su netvarkingo fono, kameros judesių ir okliuzijos problemomis net ir atsiradus formos modelių (angl. shape models) variacijom, kuriose papildomai yra atsižvelgiama į gylio informaciją (RGB+D). Antros grupės metodai pasižymėjo didesne veikimo sparta, nors tai labai priklauso nuo išankstinio aptikimo ir segmentavimo apdorojimų. Be to, jie yra priklausomi nuo vaizdo kampo pateiktuose apmokymui įvesties duomenyse, dėl ko reikalingas didesnis kiekis pavyzdžių, iš skirtingų matymo kampų, norint nustatyti tą patį veiksmą iš skirtingų apžiūros taškų [ACN⁺20].

Kita mokslininkų grupė nutarė identifikuoti svarbiausius taškus (angl. salient points) įvesties duomenų kiekviename kadre ir tokie nustatymo būdai yra žinomi kaip vietinių požymių (angl. local features) metodai [ACN⁺20].

Veiksmo atpažinimo uždavinyje svarbų vaidmenį turi ir veiksmo judesio (angl. motion) nustatymas. Dėl to buvo pasiūlytos optinio srauto (angl. optical flow) technikos, skirtos nustatyti žemo lygio vietinių požymių poslinkius bei trajektorijas iš kelių kadru, kurie galėtų būti toliau perduoti klasifikatoriams arba vėl panaudoti požymių gavyboje [ACN⁺20].

2.3. Skeleto pagrindu žmogaus pozos nustatymo sistema

Skeleto pagrindu žmogaus pozos nustatymų metodus galima suskirstyti į dvi kategorijas. Vienai kategorijai galima būtų priskirti metodus, kurie pirmiau nustato žmogaus pozos ribinį lauką (angl. bounding box) ir po to pozos taškus apibrėžtame lauke. Antros grupės metodai pirmiau nustato kiekvieną žmogaus pozos tašką, o vėliau jie sugrupuojami tokiu būdu, kad suformuotų pozą. Pirmajai grupei priklauso "AlphaPose" [FXT⁺17], "MediaPipe BlazePose" [BGR⁺20] metodai. Antrajai kategorijai priklauso "OpenPose" [CSW⁺17].

AlphaPose

Šiuo metodu galima nustatyti kaip vieno žmogaus pozą kadre, taip ir daugelių žmonių pozas. Geba nustatyti 17 žmogaus pozos taškų, kur kiekvienas taškas turi (x,y) koordinatas [FXT⁺17].

MediaPipe BlazePose

Geba nustatyti 33 žmogaus pozos taškus, kur kiekvienas taškas turi (x,y,z) koordinatas. Yra galimybė gauti kiekvieno pozos taško matomumo įvertį. Taškas pažymimas kaip nematomas ir tais atvejais kai nepavyksta nustatyti scenoje [BGR⁺20].

Svarbu paminėti, kad Z koordinatė nėra tikra gylio informacija, ji yra eksperimentinė ir dirbtinai apskaičiuota. Z ašis yra statmena kamerai ir praeina tarp žmogaus klubų. Pradžios taškas yra apytiksliai per vidurį tarp kairio ir dešiniojo klubo, ir per vidurį tarp klubų priekio ir galo kameros atžvilgiu. Teigiami Z taško įverčiai gaunami kai galūnės yra toliau nuo kameros, o neigiami kai arčiau. Z koordinatė neturi viršutinės ar apatinės ribos [BGR⁺20].

OpenPose

Geba nustatyti vieno ar daugelio žmonių scenoje pozų taškus. Žmogaus poza apibrėžiama 25 taškais, kurių kiekvienas turi (x,y) koordinatas [CSW⁺17].

2.4. Skeleto pagrindu veiksmo nustatymo sistemos

Skeleto pagrindu veikiančios sistemos pasižymi atsparumo galinio plano triukšmo problemoms, todėl yra aktyviai tyrinėjami mokslininkų ir siūlomi sprendimai, sprendžiantys kitas problemas, bei nauji pritaikymo būdai.

Viena tokių yra H.Wang, B.Yu, K.Xia, J.Li, X.Zuo mokslininkų grupė, kuri 2019 metų antroje pusėje nagrinėjo skeleto taškų gavimą iš pseudo paveikslėlių ir jų visavertį panaudojimą CNN pagrindu veikiančiose veiksmo atpažinimo sistemose. Jie pastebėjo, kad ankstesniuose metoduose nėra gerai išnaudojami kūno dalių judesio kryptis, kadangi yra naudojami vien skeleto taškų koordinatės. Todėl jie pasiūlė naują modalumo požiūrį – skeleto kraštų judesys (angl. skeleton edge motion). Remiantis šiuo požiūriu, yra apskaičiuojamas to paties judančio kūno sąnario pasukimo kampas bei atstumas tarp pirminės padėties t_1 laiku ir padėties t_2 laiku. Taip pat, jie pasiūlė naują erdvės-laiko bloką (angl. spatial-temporal block), kuris leidžia gauti efektyvesnes erdvės-laiko atvaizdus iš skeleto pseudo-paveikslėlių. Bloką sudaro praleidimo jungtys ir dvi šakos: [WYX⁺21].

1. Erdvės (angl. spatial) šaka, kurioje naudojamas $1 \times k$ konvoliucinis filtras. Šios šakos tikslas yra išmokti ryšius tarp skirtingų kūno sąnarių tame pačiame vaizdo kadre;
2. Laikinė (angl. temporal) šaka, kurioje naudojamas $k \times 1$ konvoliucinis filtras. Tikslas yra išmokti sąnario judesį kūno atžvilgių, naudojant skirtingus vaizdo kadrus [WYX⁺21].

Kiekvienos šakos rezultatas yra apjungiamas į bendrą rezultatą kartu su praleidimo jungties rezultatu ir paduodamas kaip įvestis į sekantį bloką. Savo sistemoje jie naudoja 3 tokius pasikartojančius blokus.

Rezultatų klasifikavimui yra siūloma naudoti savarankiškai prižiūrimo progresuojančios eilės nuostolio (angl. self-supervised progressive ranking loss) funkciją [WYX⁺21].

Gautos sistemos testavimui buvo panaudotas PennAction duomenų rinkinys. Jame yra 15 skirtingų veiksmų įamžintų 2326 vaizdo įrašuose. Kiekviename kadre yra sužymėta 13 sąnarių taškų. Autoriai pažymi, kad šitame duomenų rinkinyje dažnai pasireiškia sudėtingi kūno okliuzijos atvejai bei dideli dydžio skirtumai tarp objektų. Taip pat, naudoti UTD-MHAD, NTU RGB+D, NTU RGB+D 120 ir CSL duomenų rinkiniai. Eksperimento rezultatai parodė [WYX⁺21]:

1. PennAction. 93,12% tikslumas, naudojant tik RGB vaizdo įrašus. 98,71% tikslumas, panaudojus skeleto taškų informaciją iš duomenų rinkinio.
2. UTD.MHAD. 94,08% ir 95,59% tikslumai, atitinkamai naudojant tik RGB ir papildomai skeleto taškų koordinatės iš duomenų rinkinio.
3. NTU RGB+D. 80,19% ir 85,77% tikslumai, atitinkamai naudojant tik RGB ir skeleto taškų koordinatės iš duomenų rinkinio.
4. NTU RGB+D 120. 84,16% ir 85,53% tikslumai, atitinkamai naudojant tik RGB ir papildomai skeleto taškų koordinatės iš duomenų rinkinio.
5. CSL. 93,05% tikslumas.

2020 metais panašiu metu buvo pateikti vertinimui darbai iš kelių mokslininkų grupių [AJL⁺21; AKS⁺21; SP21]. T.Ahmad, L.Jin, L.Lin, G.Tang pasiūlė naudoti erdvės-laiko veiksmo

atpažinimą panaudojus grafų praretinimo (angl. graph sparsification) metodą. Kadangi visi kūno sąnariai turi skirtingą informaciją erdvės-laiko grafe, ne visi jie yra vienodai svarbūs veiksmo atpažinimui. Todėl galima sudaryti tokį grafą, kuris būtų mažesnis, kas reiškia greitesnį ir efektyvesnį tinklo modelio apmokymą. Toks metodas yra įgyvendinamas GCN architektūros pagalba su spektriniu praretinimu. Spektrinio praretinimo esmė yra sugeneruoti grafo poaibį per beveik tiesinį laiką. Tai yra padaroma pasitelkiant tikimybinį atrinkimą. Šiuo proceso metu yra ištrinamos viršūnės ir briaunos, kurios yra nustatomos kaip perteklinė informacija, todėl šitas procesas yra nuostolingas. Taip pat yra naudojamas sutelkimo mechanizmas, kuris atsižvelgia į grafo požymius ir jo topologiją. Sistemos veikimo principas susideda iš vaizdo kadru susatymo į eilę, grafų praretinimo, kelių perdavimų į grafų konvoliucijos sluoksnį bei sutelkimo mechanizmo sluoksnius bei po paskutinio konvoliucijos sluoksnio yra pritaikomas globalus dėmesio sutelkimo mechanizmas ir rezultatas yra perduodamas į pilnai sujungtą sluoksnį klasifikavimui. Susatymas į eilę vyksta remiantis erdvės-laiko požiūriu, kadangi svoriai yra skirtingi viršūnių (angl. vertices), sujungtų erdvėje, palyginus su viršūnėmis sujungtomis laike. Viršūnėms yra priskiriama didesnio svorio reikšmė, jeigu yra reikšmingas pokytis tarp nuoseklų vaizdo kadru. Grafų praretinimas remiasi grafų kraštinių mažai efektyviu atsparumu (angl. low effective resistance) bei žemos įtampos viršūnėmis (angl. low voltage vertices) analogijomis, aprašytais jų darbe [AJL⁺21].

Sistemos veikimas buvo testuojamas su vidutinio dydžio (UTD-MHAD, J-HMDB) duomenų rinkiniais, ir dideliais (NTU RGB+D, NTU RGB+D 120, Kinetics-Skeleton) duomenų rinkiniais. Eksperimentai parodė, kad: [AJL⁺21]

1. UTD-MHAD – 99,5% tikslumą.
2. J-HMDB – 87,5% tikslumą.
3. NTU RGB+D. CS – 91,3% tikslumą. CV – 96,9% tikslumą.
4. NTU RGB+D 120. CS – 67,5% tikslumą. CV – 68,8% tikslumą.
5. Kinetics-Skeleton. Top-1 36,6% tikslumą ir Top-5 54,8% tikslumą.

A.Stergio ir R.Poppe pasiūlė naują metodą pavadinimu (angl. Squeeze and Recursion Temporal Gates (SRTG)), kurio tikslas yra gauti požymius, kurie yra nuoseklus laiko atžvilgiu (angl. temporally consistent). Šis metodas padeda išspręsti problemą kai kartais apmokant 3D-CNN tinklą, gali pasitaikyti išmokti požymiai, kurie tik atsitiktinai koreliuoja su tikrai nuosekliais laiko šablonais [SP21].

SRTG metodas užtikrina, kad įvesties duomenys yra nuoseklūs laike su sumodeliuotais požymiais. Tai užtikrinama specialia laikine sklende (angl. temporal gates), kuriose tikrinama, ar cikliška pastovus metodas yra išlaikytas [SP21].

Cikliškai pastovus (angl. cyclic consistency) yra metodas, kuris palygina kadrus iš dviejų laiko sekų 1:1 principu. Tai padaroma interpretuojant plotą tarp dviejų požymių kaip įterpta erdvė (angl. embedding space) ir bandant atrasti „soft nearest neighbor“ požymį kitos laiko sekos įterptoje erdvėje. Jeigu požymiai suranda vienas kitą laiku t , vadinasi cikliškas pastovumas yra teisingas, jeigu bent vienas iš tų požymių nesuranda savo „kaimyno“, sakoma, kad įvyko laiko cikliškumo klaida (angl. temporal cyclic error) [SP21].

Pasiūlyto metodo tikrinimas buvo įvykdytas, naudojant HACS, Kinetics-700, Moments in

Time, UCF-101 ir HMDB-51 duomenų rinkinius. Metodas buvo išbandomas su skirtingo ResNet tinklo konfigūracijomis. Toliau yra pateikiami geriausi pasiekti tikslumo įverčiai, naudojant SRTG metodą [SP21]:

1. HACS. Top-1 84,326% tikslumas ir Top-5 96.852% tikslumas.
2. Kinetics-700. Top-1 56.462% tikslumas ir Top-5 77,439% tikslumas.
3. Moments in Time. Top-1 33.723% tikslumas ir Top-5 59.114% tikslumas.
4. UCF-101. Top-1 97.325% tikslumas ir Top-5 99.557% tikslumas.
5. HMDB-1. Top-1 77.536% tikslumas ir Top-5 96.253% tikslumas.

F.Afza ir jo kolegos pristatė veiksmo atpažinimo sistemą, kurioje pasiūlė išankstinį duomenų apdorojimą, taikant HSI transformacijas, silueto gavimo metodą, naudojant patobulintą EM segmentavimą sujungiant su statistine segmentacija, požymių gavimo techniką, naudojant naują ilgio valdymo apjungimo (angl. length control fusion(LCF)) būdą bei „geriausių“ požymių išrinkimo techniką [AKS⁺21].

HSI spalvų transformacijos tikslas yra pagerinti kontrastą vaizdo kadruose. Tam yra naudojami 3 kanalai (angl. channels) atspalvio, sodrumo ir intensyvumo [AKS⁺21].

Silueto gavimui buvo gauti judesio požymiai optinio srauto (angl. optical-flow) pagalba. Šie požymiai nurodo judančias dalis duotame kadre, kurios yra apskaičiuojamos naudojant kryptį nusakančius kintamuosius (u,v) ir laiko kintamąjį (t). Toliau buvo atliekamas patobulintas EM (angl. expectation-maximization (EM)) segmentavimas, kuriuo metu praretinta aktyvacijos funkcija (angl. sparse activation functions) buvo pritaikyta ir gauti pikseliai sujungiami su pikseliais, gautais pritaikius naująjį statistinį segmentavimo metodą [AKS⁺21].

Požymių gavimo žingsnyje yra gaunami požymiai trijų tipų: histograma orientuotų gradientų (angl. histogram of oriented gradients (HOG)), geometriniai ir GLCM (angl. Gray Level Co-occurrence Matrix). Tikslas yra kuo tiksliau identifikuoti veiksmą, kadangi kiekvienas jų turi skirtingą formą, tekstūrą ir atlikimo kampą. Kiekvieno tipo požymiai yra apskaičiuojami arba gaunami pritaikius atitinkamas formules bei rezultate yra pateikiami 3 vektoriai atitinkamai, kurie toliau yra apjungiami LCF būdu ir gaunama viena matrica. Kadangi gautų vektorių dydžiai skiriasi, o HOG vektorius yra žymiai didesnis už kitus 2, LCF apskaičiuoja tikimybes HOG vektoriuje ir reikšmė su didžiausia tikimybe yra parenkama ir naudojama kaip svoris kontroliuoti mažiausio ilgio vektorių [AKS⁺21].

„Geriausius“ požymiai yra pasirenkami naudojant naują pasiūlytą parametrizuotą entropijos-dispersijos (angl. Weighted Entropy-Variance) metodą. Kiekvienam požymių rinkiniui yra apskaičiuojama dispersija ir tada apskaičiuojama entropija. Pagal apibrėžtą nuostolių funkciją yra atrenkami požymiai ir formuojamas 1x500 dydžio vektorius, kuris yra naudojamas klasifikavime. Klasifikavimas yra atliekamas, naudojant M-SVM [AKS⁺21].

Pasiūlytos sistemos tikslumas buvo įvertintas naudojant Weizmann, KTH, UCF YouTube ir UCF Sports duomenų rinkinius. Per bandymus geriausi pasiekti tikslumai yra: 100% su KTH duomenų rinkiniu, 97,9% su Weizmann duomenų rinkiniu, 94,5% su UCF YouTube duomenų rinkiniu, 99,3% su UCF Sports duomenų rinkiniu [AKS⁺21].

2021 metais X.Li, Y.Hou, P.Wang, Z.Gao, M.Xu, W.Li pasiūlė naują transformerio pagrindu

(angl. transformer-based) veikiantį egocentrinio veiksmo atpažinimo karkasą (Trear). Karkasą sudaro trys moduliai: du autoenkoderiai su dėmesio sutelkimo mechanizmu paveikslėlyje (angl. inter-frame attention encoder) ir bendro dėmesio (angl. mutual-attentional) apjungimo blokas [LHW⁺21].

Kaip įvesties duomenis Trear karkasas priima RGB vaizdo kadrus ir atitinkamus gylio žemėlapius (angl. depth maps). RGB kadrai yra paduodami į vieną autoenkoderį su dėmesio sutelkimo mechanizmu paveikslėlyje, o atitinkami gylio žemėlapiai į kitą autoenkoderį. Gauti požymiai (RGB ir gylio) iš kiekvieno autoenkoderio yra perduodami į bendro dėmesio apjungimo bloką. Jis yra sudarytas iš bendro dėmesio sluoksnio ir požymio apjungimo operacijos. Bendro dėmesio sluoksnyje požymiai apsikeičia informacija, jiems yra pritaikomos kelios operacijos, praeinama atsitiktinio užmiršimo sluoksnį, liekamojo ryšio ir LayerNorm operacijas ir tada pagaminamas požymių tarp-modalinis atvaizdas (angl. cross-modality representation). Apjungti požymiai yra apdorojami tiesiniame sluoksnyje, gaunamas jų klasifikavimas ir tada perduodami galutiniam klasifikavimui [LHW⁺21].

Autoenkoderio su dėmesio sutelkimo mechanizmu paveikslėlyje veikimas. Kiekvieno kadro požymio žemėlapiams yra vykdomas sutelkimas vidurkinant su tikslu gauti įterpinių požymius. Toliau yra užkoduojama kiekvieno kadro pozicijos informacija ir veikimas tęsiamas kaip įprasto transformerio architektūroje. Vėliau vietoj rekurentinių vienetų naudojimo yra sukonstruojamos tarpkadrinės priklausomybės, naudojant „self-attention“ algoritmą. Tam kad pagerinti erdvinę koreliaciją, buvo panaudotas atsitiktinis įvesties duomenų kirpimo būdas, kurio efektyvumas buvo patvirtintas eksperimentu būdu [LHW⁺21].

Karkaso efektyvumo tikrinimui buvo panaudoti THU-READ, FPHA ir WCVS duomenų rinkiniai. Toliau yra pateikiami bandymų rezultatai [LHW⁺21]:

1. THU-READ duomenų rinkinys. Naudojant vien gylio informaciją – 76,04% tikslumas. Naudojant RGB vaizdo įrašus – 80,42% tikslumas. Naudojant kartu RGB ir gylio informaciją – 84,90% tikslumas.
2. FPHA duomenų rinkinys. Naudojant vien gylio informaciją – 92,17% tikslumas. Naudojant RGB vaizdo įrašus – 94,96% tikslumas. Naudojant kartu RGB ir gylio informaciją – 97,04% tikslumas.
3. WCVS duomenų rinkinys. Naudojant vien gylio informaciją – 63,72% tikslumas. Naudojant RGB vaizdo įrašus – 68,27% tikslumas. Naudojant kartu RGB ir gylio informaciją – 71,49% tikslumas.

2.5. Pratimų ir susijusių sistemų apžvalga

Kritimo nustatymo sistema

2020 metais E.Stone su kolegomis pasiūlė žmogaus kritimo aptikimo sistemą, kurios tikslumas yra 98,2% naudojant duomenų rinkinį sudarytą iš „UR Fall Detection“ ir „Fall Detection Dataset“ duomenų rinkinių. Šios sistemos pagrindą sudaro OpenPose skeleto taškų nustatymo sistema bei RNN LSTM/GRU modeliai. Buvo panaudotas OpenPose tikralaikio kelių žmonių 2D

pozos nustatymo algoritmas, kuris nustato 25 žmogaus sąnarių taškus ir pagal šių taškų pozicijos pasikeitimus kadrų sekoje atpažįstą judesį. Jų pasiūlytas sprendimas pasižymi efektyvumu nustatant sąnarių taškų pozicijas bei išfiltruojant supančios aplinkos triukšmą siekiant didesnio tikslumo. OpenPose algoritmas apskaičiuoja žmogaus kūno skeleto taškus naudojant sričių afininių laukų (angl. part affinity fields) algoritmą bei pasitikėjimo žemėlapi (angl. confidence map). Taip pat jie atliko duomenų išankstinį apdorojimą siekiant suvienodinti įvesties duomenų formatą. Modelio tikslumo pagerinimui buvo vykdyta realiatyvios pozicijos normalizacija (angl. relative position normalization) taikant min–maks. normalizacijos (angl. min-max normalization) metodą. Taip pat pasiūlytas normalizacijos metodas perkėlimui originalių koordinačių pozicijų į santykinės koordinačių n-otojo kadro poziciją. Tuo siekiama sumažinti nereikalingų požymių kiekį bei išvengti trūkstantų taškų apskaičiavimo poreikio. Be to, buvo pastebėta, kad kritimo judesio aptikimui užtenka 15 identifikuotų sąnarių taškų. Tam tikrose situacijose kai žmogaus pozos taškų identifikavimas negali būti atliktas teisingai, pavyzdžiui, žmogaus kūnas yra uždengtas, užtemdytas, žmogaus kūno kontūras nėra aiškus, buvo panaudotas interpoliacijos metodas. Autoriai panaudojo RNN, LSTM ir GRU modelius savo architektūroje. Buvo sukonfigūruotas 1 paslėptas sluoksniu RNN, LSTM ir GRU modeliuose ir atlikti eksperimentai [LDK⁺20].

Jogos pozų nustatymo sistema

2019 metais S.Yadav su grupe mokslininkų pristatė būdą kaip galima būtų atpažinti asanų pozas, naudojant giliuosius neuroninius tinklus iš RGB vaizdo įrašų arba iš vaizdo srauto realiuoju laiku. Jų sistemą sudaryta iš 4 žingsnių:

1. Duomenų surinkimas (realiuoju laiku arba iš išsaugotų vaizdo įrašų);
2. Pozos (sąnarių) taškų nustatymas iš pateiktų duomenų;
3. Šablonų nustatymas naudojant CNN modelį, kurių analizę atsižvelgiant į laiką, atlieka LSTM;
4. Prognozių apskaičiavimas ir rezultato pateikimas.

Antrame žingsnyje taikomas OpenPose algoritmas kūno taškų identifikavimui, naudojant dalinius giminingus laukus ir pasitikėjimo žemėlapius. Gautų taškų koordinatės yra užrašomos į JSON tipo failą. Toliau laike paskirstytame CNN sluoksnyje yra gaunami požymių žemėlapiai iš taškų koordinačių. Yra naudojama 16 filtrų 3x3 dydžio ir ReLU aktyvacijos funkcija pritaikoma kiekvienam vaizdo kadrai. Taip pat, CNN sluoksniu rezultatui yra pritaikoma rinkinio normalizacija (angl. batch normalization), kad paspartinti apmokymą. Toliau yra išmetimo sluoksniu. Galiausiai kiekvienam 45-am kadro CNN sluoksniu rezultatas yra ištiesinamas ir perduodamas LSTM modeliui. LSTM modelyje, sudarytame iš 20 vienetų (angl. unit), yra atpažįstami laiko požymių pokyčiai ir visas pratimas nuo pradžios iki jo užbaigimo laikomas vientisa atpažinta veikla. LSTM modelio rezultatas yra perduodamas Softmax algoritmui, kuris nustato asanų klasių tikimybės [YSG⁺19].

Atlikus sistemos veikimo realiuoju laiku bandymus, buvo identifikuota problema, kad jeigu vertinimui yra naudojamas vienas kadras, bendras sistemos tikslumas krenta iki 60%. Ši problema buvo išspręsta sutelkiant (angl. pooling) 45 kadrų informaciją (atitinka 1,5 sekundės, nes įrašymo

greitis 30 kadrų per sekundę). Papildomai tai padėjo išspręsti problemas kai žmogus tik pradeda atlikinėti pratimą arba jį jau užbaiginėja (saugiai grįžta iš tos pozicijos). Tai yra svarbu, nes „įėjimas“ į pozą ir išėjimas iš jos yra pratimo dalis ir turi būti atlikti teisingai [YSG⁺19].

Taip pat, jie sukūrė duomenų rinkinį, kuriame buvo nufilmuotos 6 asanų pozos atliekamos skirtingų 15 žmonių (10 vyrų ir 5 moterų). Žmonės atlikinėjo asanų pozas priešais kamerą, nutolę per 4-5 metrus nuo jos. Duomenų rinkinį sudaro RGB formato vaizdo įrašai. Vieno įrašo trukmė yra apie 45 sekundės. Filmavimui buvo pasirinktas 30 kadrų per sekundę greitis [YSG⁺19].

Sistema parodė 99,04% tikslumą naudojant vieną kadrą pozos nustatymui ir 99,38% tikslumą, naudojant sutelktus 45 kadrus iš testavimui skirtų vaizdo įrašų. Išbandžius sistemą realiomis veikimo sąlygomis buvo deklaruotas 98,92% tikslumas. Tuose bandymuose dalyvavo kiti 12 žmonių (5 vyrai ir 7 moterys) negu apmokymui skirtuose vaizdo įrašuose [YSG⁺19].

2.6. Daugiatikslio uždavinio optimizavimas

Daugiatiksliai kompiuterinės regos sprendimai medicinoje gali sumažinti gydytojų krūvį, suteikiant galimybę pacientams savarankiškai atlikti pratimus ir jų vertinimus [DBZ23]. Tokie sprendimai reikalauja vienu metu atsižvelgti į kelis skirtingus veiksnio vertinimo tikslus, todėl būtinas daugiatiksliasis metodas. Tai reiškia, kad reikia ne tik optimizuoti atskirą funkciją, bet ir ieškoti pusiausvyros tarp įvairių funkcijų, kurios dažnai yra tarpusavyje priešingos: gerinant vienos funkcijos rezultatus, gali pablogėti kiti. Optimizavimo procese svarbu atrasti Pareto optimalius sprendinius, kurie užtikrintų geriausią įmanomą visų funkcijų derinį [Lan13]. Sprendinys x^* yra Pareto optimalus, jei nėra kito sprendinio x , kuris tenkintų sąlygą:

$$f_i(x) \leq f_i(x^*), \forall i \in \{1, \dots, k\}$$

ir bent vienam j galioja, kad:

$$f_j(x) < f_j(x^*)$$

Aibė funkcijų reikšmių, kurios sudaro Pareto optimalius sprendinius, vadinama Pareto frontu.

3. Veiksmų atpažinimo duomenų rinkiniai

Šiame skyriuje yra apžvelgiami duomenų rinkiniai, kurie savo sudėtyje gali turėti duomenų ar jų dalių, tinkančių apmokyti neuroninį tinklą identifikuoti pratimus, paminėtus Tinetti teste (žr. priedas 1).

KTH duomenų rinkinys

Sukurtas 2004 metais. Pateikiami 6 veiksmai (vaikščiojimas, bėgiojimas, bėgimas, plojimas rankomis, mojavimas ranka ir bokso užsiėmimas). Filmavimai vyko 4 skirtinguose kontroliuojamose aplinkose kaip viduje taip ir lauke. Veiksmus atlikinėjo 25 žmonės. Kameros priartinimo galimybė buvo naudota kameros judėjimo simuliacijai. Bendras duomenų rinkinio dydis yra 1,147 TB [SLC04].

NTU RGB duomenų rinkiniai

NTU RGB+D [SLN⁺16]. Šiame duomenų rinkinyje yra pateikta 56880 vaizdo įrašų, kurie perteikia 60 skirtingų veiksmų (klasių). Visus veiksmus galima suskirstyti į 3 dideles grupes: 1) kasdieniai veiksmai 2) medicininių simptomų veiksmai 3) dviejų žmonių sąveikos veiksmai.

NTU RGB+D 120 [LSP⁺19]. Šitas duomenų rinkinys apima NTU RGB+D duomenų rinkinio duomenis bei papildomas 60 veiksmų klases, kurios buvo užfiksuotos naujuose 57600 vaizdo įrašuose ir šiuo metu yra didžiausias žmogaus veiksmų atpažinimo mokymui naudojamas duomenų rinkinys. Bendrai duomenų rinkinyje yra pateikiama 120 veiksmo klasių ir 114480 vaizdo įrašų, kuriems įrašyti vaidino 106 žmonės skirtingose vietose.

Abiejuose duomenų rinkiniuose yra pateikiami RGB vaizdo įrašai, gylio žemėlapių sekos (angl. depth map sequences), 3D skeleto duomenys, infraraudonųjų spindulių (IR) vaizdo įrašai kiekvienam pavyzdžiui. Filmavimas vyko vienu metu naudojant 3 Kinect V2 kameras. RGB vaizdo įrašų rezoliucija yra 1920x1080 pikselių. Gylio žemėlapių ir IR vaizdo įrašų raiška yra 512x424 pikselių. 3D skeleto duomenyse yra pateikiamos trimatės (3D) koordinatės 25 kūno sąnarių (angl. body joints), esančių kiekviename kadre (angl. frame). Pilno šio duomenų rinkinio dydis yra 2,3 TB. Iš kurių: 3D skeletų duomenų dydis yra 10,3 GB, užmaskuotų gylio žemėlapių (angl. masked depth maps) dydis yra 147 GB, pilno gylio žemėlapių (angl. full depth maps) dydis yra 1,435 TB, RGB vaizdo įrašų dydis yra 260 GB, IR vaizdo įrašų dydis yra 389 GB.

UTD-MHAD duomenų rinkinys

Apima 27 skirtingų veiksmų sąrašą, tarp kurių yra „bėgiojimas vietoje“, „ėjimas vietoje“, „atsistojimas iš sėdimos pozicijos“, „atsisėdimas iš stovimos pozicijos“, „priekinis svyravimas“ bei „pritūpimai“. Vaizdo įrašai buvo nufilmuoti Kinect kameros pagalba. RGB rezoliucija yra 640x480 pikselių, gylio sensorius 16-bitų raiškos 320x240 pikselių. Filmavimuose dalyvavo 8 žmonės (4 vyrai ir 4 moterys). Duomenų rinkinyje yra pateiktos RGB vaizdo įrašai, gylio informacija, skeleto taškų informacija bei inercinio jutiklio signalai. RGB – 1,03 GB, gylio informacija – 121 MB,

skeleto taškų 14,3 MB, jutiklio signalų dydis – 5,10 MB. Bendras dydis – apytiksliai 1,17 GB [C]K15].

Florence 3D Actions duomenų rinkinys

Sukurtas 2012 metais naudojant Kinect kamera. Pateikiami 9 veiksmų tipai (mojavimas, gėrimas iš butelio, atsiliepimas į skambutį, plojimas, užsirišti raištelį, atsisėdimas, atsistojimas, žiūrėjimas į laikrodį, pagarbus nusilenkimas (angl. bow), kurie yra atliekami 10 skirtingų aktorių. Veiksmai atliekami esant vienam aktoriui scenoje arba keliems žmonėms. Vaizdo įrašai yra nufilmuoti 640x480 rezoliucijos 20 kadrų per sekundę greičiu. Iš viso yra 215 vaizdo įrašų, iš kurių 20 skirta atsisėdimui ir 20 atsistojimui. Bendras duomenų rinkinio dydis yra 324 MB [SVB⁺13].

MSR DailyActivity3D duomenų rinkinys

Duomenų rinkinys sukurtas 2012 metais naudojant Kinect kamera. Jį sudaro gilio informacijos rinkmenos, žmogaus pozos taškų koordinatų rinkmenos ir vaizdo įrašai, demonstruojantys 16 skirtingų veiksmų: gėrimą, valgymą, knygos skaitymą, skambinimą mobiliuoju telefonu, rašymą ant popieriaus lapo, naudojimąsi nešiojamu kompiuteriu, naudojimąsi dulkių siurbliu, linksminimąsi, ramų sėdėjimą, popieriaus mėtymą, žaidimą su žaidimo konsolės valdikliu, gulėjimą ant sofos, vaikščiojimą, grojimą gitara, atsistojimą, atsisėdimą. Dalyvavo 10 aktorių. Scenoje yra sofa, fone matomi 2 darbo kambariai (ofiso tipo), kuriuose kartais matomos vieno ar dviejų žmonių galvos. Iš viso yra 320 vaizdo įrašų, iš kurių, po 20 vaizdo įrašų skirta ramiam sėdėjimui ant sofos, atsistojimui, atsisėdimui. Vaizdo įrašai yra nufilmuoti 640x480 rezoliucijos, 30 kadrų per sekundę greičiu. Bendras duomenų rinkinio dydis yra 24,5 GB [WLW⁺12].

4. Tinetti duomenų rinkinys

Filmavimo strategija

Pasitarus su kineziterapeutu buvo apibrėžta kaip turi būti tinkamai filmuojami Tinetti metodika vertinami pusiausvyros uždaviniai, kad būtų įmanoma įvertinti atliekamų uždavinių teisingumą. Ši strategija apima uždavinių filmavimo kampą, filmuojamų objektų atstumą iki kameros, uždavinių filmavimo trukmę.

Atstumas iki kameros.

Visos užduotys turi būti filmuojamos pakankamu atstumu nuo kameros, kad būtų matomas žmogus pilnu ūgiu, jo kojų pozicija.

Filmavimo kampas.

Tinetti-1, Tinetti-2 ir Tinetti-9 užduotys turėtų būti filmuojamos 45° laipsnių kampu. Tikslas, kad būtų matoma nugaros linija, pėdų bei rankų pozicijos. Tinetti-3, Tinetti-4, Tinetti-5, Tinetti-8 gali būti filmuojami iš priekio, kadangi atsižvelgiama į užduoties bandymų įvykdyti kiekį, kojų atramos plotį ir pagalbinių priemonių naudojimą.

Filmavimo trukmė/užduoties trukmė.

Tinetti-3 ne mažesnis negu 5 sekundės įskaičiuojant laiką atsistojimą nuo kėdės. Visoms kitoms užduotims tiek, kiek reikia žmogui. Dažniausiai teisingam atlikimui užtenka nuo kelių iki 10-15 sekundžių. Neteisingi bandymai gali užtrukti pora minučių.

4.1. Sukurtas duomenų rinkinys

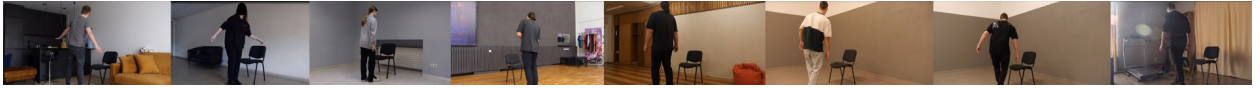
Yra nufilmuotos Tinetti-1, Tinetti-2, Tinetti-8 ir Tinetti-9 užduotys. Iš viso gautas 581 vaizdo įrašas, iš kurių 206 su teisingai atliekamais ir 375 su neteisingai atliekamais pratimais. Žiūrėti vaizdo įrašų pasiskirstymą lentelėje 2.

2 lentelė. „Tinetti“ pusiausvyros vertinimo užduočių vaizdo medžiagos detalizavimas

Unikalus numeris	Pavadinimas	Teisingi (vaizdo įrašų skaičius)	Neteisingi (vaizdo įrašų skaičius)
Tinetti-1	Pusiausvyros laikymas sėdint	39	59
Tinetti-2	Atsistojimas iš sėdimos padėties	50	127
Tinetti-8	Apsisukimas 360° kampu	48	89
Tinetti-9	Atsisėdimas	69	100
Iš viso:		206	375

Vaizdo įrašai yra nufilmuoti 1920x1080 rezoliucija, 25 kadrai per sekundę greičiu.

Vaizdo įrašai yra nufilmuoti 7 skirtingose vietose (žr. paveikslėlį 7), iš skirtingų kampu ir pakankamu atstumu nuo aktoriaus. Apšvietimas scenose skiriasi, 6 vietose yra skirtingas dirbtinio šaltinio apšvietimas, 1 vietoje filmuota su netiesioginė dienos šviesa. Taip pat, filmavimo scenose skiriasi fonas, aplinkui filmuojama aktorių yra pašalinių (nejudančių) objektų.



7 pav. Sukurto duomenų rinkinio filmavimo aplinkų apšvietimo pavyzdžiai

Filmavimuose dalyvavo 11 asmenų, 4 moteriškos lyties ir 7 vyriškos lyties. Dalyviai yra skirtingo amžiaus, ūgio, kūno sudėjimo, fizinių pajėgumų, apsirengę sportinę aprangą arba laisvalaikio tipo drabužiais (džinsai ir marškinėliai).

„Tinetti“ duomenų rinkinys surūšiuotas pagal žmogaus taškų matomumo įvertį, atlikimo teisingumą ir veiksmo klasę. Duomenų rinkinyje nėra pateikiami originalūs vaizdo įrašai, tik MediaPipe BlazePose (žr. skryrius 2.3) pagalba gauti žmogaus pozos taškai, kurie yra išsaugoti rinkmenoje su plėtiniu „.csv“. Duomenų rinkinys turi 4 atskirus rinkinius:

1. Žmogaus taškai teisingai atliekamų pratimų su matomumo įverčiais „*Keypoints_True_Full*“
2. Žmogaus taškai neteisingai atliekamų pratimų su matomumo įverčiais „*Keypoints_False_Full*“
3. Žmogaus taškai teisingai atliekamų pratimų be matomumo įverčių „*Keypoints_True_NoVis*“
4. Žmogaus taškai neteisingai atliekamų pratimų be matomumo įverčių „*Keypoints_False_NoVis*“

Kiekvienas „*Keypoints_X_Y*“ rinkinys turi atskirus keturis aplankus atitinkančius Tinetti pratimus „360Turnaround“ (Tinetti-8), „BalancingWhileSitting“ (Tinetti-1), „Sit“ (Tinetti-9), „Standup“ (Tinetti-2). Kiekviename pratimo aplanke yra rinkmenos su plėtiniu „.csv“, kur kiekviena rinkmena atitinka (pagal pavadinimą) vieną apdorotą vaizdo įrašą. Viename vaizdo įrašė yra nufilmuotas tik 1 žmogus atliekantis 1 tipo (teisingą arba neteisingą) Tinetti pratimą. „.csv“ rinkmenoje yra išsaugota žmogaus taškų informacija toliau išvardintu pavidalu. Pirmą eilutę – antraštę, likusios eilutės žmogaus taškų informacija. Viena informacijos eilutė atitinka viename vaizdo įrašo kadre nustatytą žmogaus taškų informaciją. Antraštės formatas, atitinka informacijos eilutės stulpelių skaičių, yra vienas iš nurodytų:

1. Rinkmenoms be taškų matomumo įverčių R , kur pt yra žmogaus pozos taškas

$$R = \{(pt_{1x}, pt_{1y}, pt_{1z}), (pt_{2x}, pt_{2y}, pt_{2z}), \dots, (pt_{Nx}, pt_{Ny}, pt_{Nz}), P, K\}, N = 33, \quad (21)$$

P – *pratimas*, $P \in \{0, 1, 2, 3\}$, K – korektiškumas, $K \in \{0, 1\}$

2. Rinkmenoms su taškų matomumo įverčiais R_{vis} , kur pt yra žmogaus pozos taškas

$$R_{vis} = \{(pt_{1x}, pt_{1y}, pt_{1z}, t_{1vis}), (pt_{2x}, pt_{2y}, pt_{2z}, pt_{2vis}), \dots, (pt_{Nx}, pt_{Ny}, pt_{Nz}, pt_{Nvis}), P, K\}, N = 33, \quad (22)$$

P – *pratimas*, $P \in \{0, 1, 2, 3\}$, K – korektiškumas, $K \in \{0, 1\}$

Pratimo reikšmių koduotė atitinka 0 - "Sit", 1 - "Standup", 2 - "BalancingWhileSitting", 3 - "360Turnaround". Korektiškumo 0 - neteisingas, 1 - teisingas.

4.2. Duomenų rinkinio praplėtimas

Praplėtimui naudoti Florence 3D Actions ir MSR DailyActivity3D duomenų rinkiniai. Pavyko atrinkti 72 vaizdo įrašus. Dalis atrinktų vaizdo įrašų buvo priskirti prie teisingai atliekamų, o kita dalis prie neteisingai atliekamų pratimų. Atrinkus vaizdo įrašus ir juos suskirsčius į kategorijas, buvo pritaikytas MediaPipe BlazePose žmogaus taškų gavimo metodas.

NTU RGB duomenų rinkiniai nėra naudojami dėl apribojimų nurodytų naudojimo taisyklėse.

Daugiau gauto duomenų rinkinio praplėtimų detalių yra pateikta 3 lentelėje.

3 lentelė. Gauto duomenų rinkinio praplėtimo detalizavimas

Unikalus numeris	Pavadinimas	Teisingi (vaizdo įrašų skaičius)	Neteisingi (vaizdo įrašų skaičius)
Tinetti-1	Pusiausvyros laikymas sėdint	9	0
Tinetti-2	Atsistojimas iš sėdimos padėties	19	14
Tinetti-8	Apsisukimas 360° kampu	0	0
Tinetti-9	Atsisėdimas	21	9
	Iš viso:	49	23

Duomenų rinkinio praplėtimas atitinka sukurto Tinetti duomenų rinkinio struktūrą, aprašyta skyriuje 4.1.

5. Eksperimentai

5.1. Duomenų paruošimas

Kiekvienam eksperimentui nuskaitant pradinį Tinetti duomenų rinkinį, yra pritaikomos transformacijos (jeigu nenurodyta kitaip):

1. Duomenų agregavimas. M - iš eilės einančių kadro taškų perkėlimas į atitinkamą skaičių stulpelių. Tokiu būdu mes suformuojam apmokymui skirtą rinkinį, kurio kiekvienoje eilutėje kūno pozos taškas pt užrašytas $(pt_{1_{x_1}}, pt_{1_{y_1}}, pt_{1_{z_1}}), \dots, (pt_{1_{x_M}}, pt_{1_{y_M}}, pt_{1_{z_M}}), M \in \mathbb{N}$
 2. Agreguotų duomenų stulpeliai yra surikiuojami tokiu būdu, kad kiekvienos to paties tipo žmogaus taško X koordinatės išsidėstytų iš eilės, Y iš eilės, Z iš eilės ir kitiems taškams kartojama tas pats. Formaliai, kai kūno pozos taškas pt , surikiuota eilutė atrodo $(pt_{1_{x_1}}, \dots, pt_{1_{x_M}}), (pt_{1_{y_1}}, \dots, pt_{1_{y_M}}), (pt_{1_{z_1}}, \dots, pt_{1_{z_M}}),$
 $\dots,$
 $(pt_{N_{x_1}}, \dots, pt_{N_{x_M}}), (pt_{N_{y_1}}, \dots, pt_{N_{y_M}}), (pt_{N_{z_1}}, \dots, pt_{N_{z_M}}), kur N = 33, M \in \mathbb{N}$
- Tai leidžia išgauti taško koordinatės pokytį kas M kadru, iš esmės, nuoseklų pokytį laike.

Duomenų įvestis yra žmogaus pozos taškų koordinatės. Duomenų rinkinio be matomumo įverčių įvesties dimencija yra 990, kas atitinka 33 taškų (x,y,z) koordinatės iš 10 vaizdo kadru. Duomenų rinkinio su matomumo įverčiais įvesties dimencija yra 1320, kas atitinka 33 taškų (x,y,z) koordinatės ir taško matomumo įvertį iš 10 vaizdo kadru.

Duomenų rinkinys yra suskirstomas į apmokymo ir testavimo pogrupius remiantis 60:40 proporcija. Yra užtikrinama, kad žmogaus taškų informacija iš vieno vaizdo įrašo nėra vienu metu ir apmokymo, ir testavimo pogrupiuose. Toliau yra atliekamas duomenų maišymas pogrupiuose.

5.2. Tyrimo aplinka

Kaip eksperimentų aplinka buvo pasirinktas "Google Colab". Toliau yra išvardijamos pagrindinių naudojamų bibliotekų versijos.

- "Python" 3.10
- "Keras" API sąsaja "Tensorflow" bibliotekos 2.15.0 versijos pagrindu
- "MediaPipe" biblioteka 0.10.11 versija
- "Pandas" biblioteka 2.0.3 versijos
- "OApacakage" biblioteka 2.7.13 versijos

Aplinkos sistema turi 12GB RAM. Apmokymams naudojama grafine vaizdo plokštė NVIDIA Tesla T4, turinti 16GB RAM.

5.3. Modelio pasirinkimas

Iš pradžių buvo iškelta hipotezė, kad naujai sudaryto duomenų rinkinio eksperimentams bus naudojamas minimas „Jogos pozų nustatymo sistemos“ modelis, sprendžiantis panašų klasifikavimo užduotį. [YSG⁺19] modelyje kaip įvesties duomenys naudojami požymiai gauti iš paveikslėlių

kadru. Po pirmųjų bandomųjų eksperimentų, paaiškėjo, kad modelis blogai sprendžia šio darbo klasifikavimo užduotis naudojant naujai sukurtą Tinetti duomenų rinkinį.

Tada buvo iškelta hipotezė, modelio apmokymui ir testavimui naudoti iš vaizdo įrašų kadru gautus žmogaus pozos taškus, pritaikius tam tikras transformacijas. Bandyto keliu buvo apibrėžti 5 CNN modeliai (bei kelios jų variacijos) dviejų klasių klasifikavimo bei daugelio klasių klasifikavimo užduotims spręsti. Daugiatikslio klasifikavimo užduočiai bandymo būdu yra apibrėžti 5 modeliai pasirinkus dviejų/daugelio klasių klasifikavimo modeliais.

5.4. Modelių architektūros

5.4.1. M2

Modelis M2, lentelė 4, yra sudarytas iš dviejų konvoliucijos, dviejų sutelkimo imant maksimalią reikšmę sluoksnių, trijų rinkinio normalizacijos, vieno ištiesinimo ir tankinimo sluoksnių. Kaip aktyvacijos funkcija yra naudojama dalimis tiesinė funkcija (Relu). Paskutiniame sluoksnyje yra pritaikoma eksponentinio normalizavimo Softmax funkcija į 4 arba 2 galimus rezultatus. Dviejų klasių klasifikavimo uždavinio atveju, iš viso yra 2022850 parametrų (7.72 MB), iš jų apmokamų - 2022402 (7.71 MB), neapmokamų - 448 (1.75 KB). Daugelio klasių klasifikavimo atveju, iš viso 2023108 parametrų (7.72 MB), 2022660 - apmokamų (7.72 MB), 448 - neapmokamų (1.75 KB).

4 lentelė. M2 modelis

Transformacija	Dimencija	Filtrų skaičius	Filtro dydis	Žingsnis	Aktyvacijos funk.	Parametrai
Input	(990, 1)					
Convolutional	(988, 32)	32	3	1	Relu	128
Batch normalization	(988, 32)	-	-	-	-	128
Max pooling	(494, 32)	-	-	2	-	0
Convolutional	(492, 64)	64	3	1	Relu	6208
Batch normalization	(492, 64)	-	-	-	-	256
Max pooling	(246, 64)	-	-	2	-	0
Flatten	(15744)	-	-	-	-	0
Dense	(128)	-	-	-	Relu	2015360
Batch normalization	(128)	-	-	-	-	512
Dense	(4)	-	-	-	Softmax	516

5.4.2. M3

Modelis M3 yra modelio M2 variacija, panaudojus ELU aktyvacijos funkciją vietoj RELU. Modelio parametrų skaičius atitinka aprašyto M2 modelio skaičių.

5.4.3. M4

Modelis M4 yra modelio M2 variacija, panaudojus GELU aktyvacijos funkciją vietoj RELU. Modelio parametrų skaičius atitinka aprašyto M2 modelio skaičių.

5.4.4. M5

Modelis 5, lentelė 22, yra sudarytas modelio M2, skyrius 5.4.1, pagrindu, pridėjus išmėtimo sluoksnį po kiekvieno sutelkimo imant maksimalią reikšmę sluoksnio (angl. max pooling) ir po paskutinio rinkinio normalizavimo sluoksnio. Modelio parametrų skaičius atitinka aprašyto M2 modelio skaičių.

5.4.5. Kiti modeliai

M6

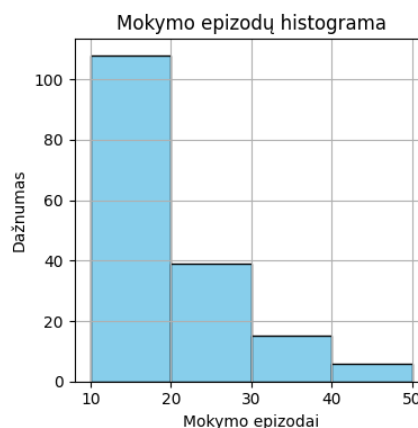
M6 modelio apibendrinimas pateiktas Priede 2 lentelėje 23. Dviejų klasių klasifikavimo uždavinio atveju iš viso turi 15810 parametrų (61.76 KB), iš jų apmokamų - 15362 (60.01 KB), neapmokamų - 448 (1.75 KB).

M7

M7 modelio apibendrinimas pateiktas Priede 2 lentelėje 24. Dviejų klasių klasifikavimo uždavinio atveju iš viso turi 1164162 parametrų (4.44 MB), iš jų apmokamų - 1164162 (4.44 MB), neapmokamų - 0.

5.5. Mokymo epizodų pasirinkimas

Iš atliktų eksperimentų mokymo epizodų statistikos (žr. pav. 8) matyti, kad didžioji dalis eksperimentų pasibaigė anksčiau negu nurodytas maksimalus mokymų epizodų skaičius. Iš to darytina išvada, kad atlikinėti eksperimentus su daugiau leidžiamų mokymo epizodų skaičiumi nėra prasmės. Todėl tolimesniuose eksperimentuose yra naudojamas didžiausias leidžiamas mokymo epizodų skaičius - 50.



8 pav. Mokymo epizodų pasiskirstymas

5.6. Klasifikavimo slenkstinių reikšmių paieška

Klasifikavimo slenkstinių reikšmių paieškos metu yra nustatomos optimalios ribinės reikšmės su tikslu sumažinti tikimybinį pagrindų veikiančio klasifikavimo modelio neteisingų teigiamų

prognozių arba neteisingų neigiamų prognozių skaičių. Šie atvejai gali būti identifikuoti, apskaičiuavus preciziškumo (angl. precision), formulė 23, ir jautrumo (angl. recall), formulė 24, metrikas.

Preciziškumas (angl. precision) - parodo santykį kiek teisingų prognozių yra iš visų teigiamų prognozių.

Jautrumas (angl. recall) - parodo kiek gerai modelis geba atlikti teisingas prognozes.

$$\text{Preciziškumas} = \frac{TP}{(TP + FP)} \quad (23) \quad \text{Jautrumas} = \frac{TP}{(TP + FN)} \quad (24)$$

TP - teisinga teigiama prognozė

TN - teisinga neigiama prognozė

FP - neteisinga teigiama prognozė

FN - neteisinga neigiama prognozė

Preciziškumas yra svarbus tuo metu kai svarbu turėti teisingas teigiamas prognozes. Jautrumas svarbu tada kai teikiamas prioritetas nustatyti visas teigiamas reikšmes ir neteisingos teigiamos prognozės yra toleruotinos.

Nagrinėjamų darbo uždavinių atveju yra svarbu kaip teisingai nustatytas pratimas, taip ir teisingai nustatytas jo atlikimo teisingumas. Todėl nuspręsta darbe modelio vertinime vadovautis F1 metriką, formulė 25, kadangi F_1 yra harmoningas vidurkis tarp preciziškumo ir jautrumo.

$$F_1 = 2 \times \frac{\text{preciziškumas} \times \text{jautrumas}}{\text{preciziškumas} + \text{jautrumas}} \quad (25)$$

Optimizavimo žingsniai

1. Paruošiama duomenų rinkinio struktūra DR su etalono reikšmėmis ir modelio prognozėmis.
2. Tada DR yra surūšiuojamas pagal vaizdo įrašų identifikatorius ir vaizdo įrašų N-agreguotų kadrų eilučių eilės numerius.
3. Tada su kiekviena ribine reikšme iš intervalo $[0..1]$ su žingsniu 0.01 yra:
 - 3.1. apskaičiuojama optimali ribinė reikšmė, palyginant DR prognozę su ribine reikšme ir jeigu prognozė daugiau už ją, rezultatas yra išsaugomas kaip 1, priešingu atveju 0.
 - 3.2. išsaugoti lyginimo rezultatai žingsnyje a) yra palyginami su DR etalono reikšmėmis ir atitinkamai rezultatas yra išsaugomas į naują stulpelį.
 - 3.3. DR yra sugrupuojamas pagal vaizdo įrašų identifikatorius.
 - 3.4. kiekvienos vaizdo grupės rezultato vidurkis apskaičiuojamas ir tada gaunamas viso DR vidurkio rezultatas.
 - 3.5. apskaičiuota optimali ribinė reikšmė su vidurkio rezultatu yra išsaugomi į ribinių reikšmių ir vidurkių rinkinį.

- Ribinių reikšmių ir vidurkių rinkinyje randama mažiausia optimali ribinė reikšmė su didžiausiu vidurkiu.

5.7. Dviejų klasių klasifikavimo eksperimentai

5.7.1. Tikslas

Šių eksperimentų tikslas yra apmokyti modelį nustatyti ar pasirinktas Tinetti pratimas yra atliekamas teisingai arba ne. Mokymai yra atskirai atliekami Tinetti-1, Tinetti-2, Tinetti-8, Tinetti-9 pratimams.

5.7.2. Konfigūracija

Pasirinkta optimizatoriaus funkcija - Adam. Pasirinkta nuostolių funkcija - kryžminė entropija. Daugiausiai leidžiama 50 apmokymo epizodų. Mokama dinaminį skaičių epizodų stebint nuostolio funkcijos validavimą ir tikrinant su 10 epizodų uždelsimu (angl. patience) iki kada validavimo nuostolis mažėja. Apmokama pasirinkus atskirai $DR_s = [4, 8, 16, 32, 64, 128]$ duomenų rinkinio dydžius (angl. batch size).

5.7.3. Rezultatai

Yra atlikta 464 eksperimentų, 168 naudojant duomenų rinkinį be taškų matomumo duomenų, 168 su taškų matomumo įverčiais ir 128 naudojant duomenų rinkinio praplėtimą su taškų matomumo įverčiais. Lentelėje 5 yra pateikiami tarp visų eksperimentų vidutiniškai 3-ų geriausių modelių rezultatai. Iš jų matoma, kad M2 modelis pasiekė vidutiniškai aukščiausią F1 statistiką, 72.15% ir 73.46% atitinkamai naudojant duomenų rinkinius be matomumo ir su matomumo įverčiais. Duomenų rinkinio su praplėtimu ir matomumo įverčiais atveju, vidutiniškai aukščiausią F1 statistiką rodo M5 modelis, 68.11%.

5 lentelė. Vidutiniškai 3-ų geriausių modelių rezultatai naudojant apmokymui duomenis be žmogaus pozos taškų matomumo, su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais

Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	72.15%	0.3472	73.46%	0.35681	67.45%	0.53253
M3	71.74%	0.4135	72.16%	0.35369	62.11%	0.73832
M5	71.60%	0.3088	70.88%	0.32681	68.11%	0.40195

Rezultatai pagal atskirus pratimus

Tinetti-1

Lentelėje 6 yra pateikti Tinetti-1 pratimo 3-ių geriausių modelių suvidurkinti rezultatai kai modelių apmokymui naudoti duomenų rinkiniai: be žmogaus pozos taškų matomumo įverčių,

su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais. Iš rezultatų palyginimo lentelės matoma, kad vidutiniškai aukščiausios F1 statistikos pasiektos:

1. M3 modelis 73.07% - duomenų rinkinio be matomumo įverčių atveju
2. M1 modelis - 75.11% - duomenų rinkinio su matomumo įverčiais atveju
3. M1 modelis - 81.77% - duomenų rinkinio su praplėtimu ir matomumo įverčiais atveju

6 lentelė. Tinetti-1 pratimo 3-ų geriausių modelių suvidurkinti rezultatai apmokant naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių, naudojant duomenų rinkinį su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais palyginimas

Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	72.90%	0.6593	75.11%	0.63263	81.77%	0.76249
M3	73.07%	0.7607	69.32%	0.7102	67.65%	1.38641
M5	69.80%	0.5572	66.34%	0.63704	80.23%	0.44842

Tinetti-2

Lentelėje 7 yra pateikti Tinetti-2 pratimo 4-ių geriausių modelių suvidurkinti rezultatai, modelių apmokymui naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių, duomenų rinkinį su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais. Iš eksperimento rezultatų matoma, kad vidutiniškai aukščiausios F1 statistikos yra: M4 modelio - 64.82%, M2 modelio - 64.92% ir M5 modelio - 51.49%, atitinkamai duomenų rinkinių, be matomumo įverčių, su matomumo įverčiais ir su praplėtimu ir matomumo įverčiais, atvejais.

7 lentelė. Tinetti-2 pratimo 4-ų geriausių modelių suvidurkinti rezultatai apmokymui naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių, naudojant duomenų rinkinį su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais palyginimas

Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	60.87%	0.26334	64.92%	0.22473	47.49%	0.43118
M3	62.73%	0.22057	64.19%	0.21742	45.29%	0.39404
M4	64.82%	0.21675	63.04%	0.20688	47.23%	0.51904
M5	64.09%	0.19706	60.86%	0.21743	51.49%	0.39366

Tinetti-8

Lentelėje 8 yra pateikti Tinetti-8 pratimo 4-ių geriausių modelių suvidurkinti rezultatai. Šių modelių apmokymui naudotas duomenų rinkinys be žmogaus pozos taškų matomumo įverčių ir duomenų rinkinys su matomumo įverčiais. Eksperimentai su praplėtimu ir matomumo įverčiais

nėra atlikti, nes Tinetti-8 pratimas nėra praplėstas vaizdo įrašais iš praplėtimui naudotų duomenų rinkinių. Iš rezultatų matoma, kad M3 modelis turi vidutiniškai aukščiausią F1 statistiką abiejų duomenų rinkinių eksperimentų atvejais, 73.08% be matomumo įverčių ir 72.62% su matomumo įverčiais.

8 lentelė. Tinetti-8 pratimo 4-ų geriausių modelių suvidurkinti rezultatai apmokymui naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių ir naudojant duomenų rinkinį su matomumo įverčiais palyginimas

Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	71.11%	0.25419	72.27%	0.23954	-	-
M3	73.08%	0.25305	72.62%	0.24499	-	-
M4	72.54%	0.24218	71.45%	0.30891	-	-
M5	68.32%	0.25951	71.71%	0.23728	-	-

Tinetti-9

Lentelėje 9 yra pateikti Tinetti-9 pratimo 4-ių geriausių modelių suvidurkinti rezultatai, modelių apmokymui naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių, duomenų rinkinį su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais. Aukščiausią F1 statistiką rodo M5 modelis, duomenų rinkinių be praplėtimo be/su matomumo įverčių atvejais, atitinkamai 84.18% ir 84.59%. Duomenų rinkinio su praplėtimu ir matomumo įverčiais atveju M3 modelis parodo vidutiniškai aukščiausią, 73.38%, F1 statistiką tarp visų modelių.

9 lentelė. Tinetti-9 pratimo 4-ų geriausių modelių suvidurkinti rezultatai apmokymui naudojant duomenų rinkinį be žmogaus pozos taškų matomumo įverčių, naudojant duomenų rinkinį su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu ir matomumo įverčiais palyginimas

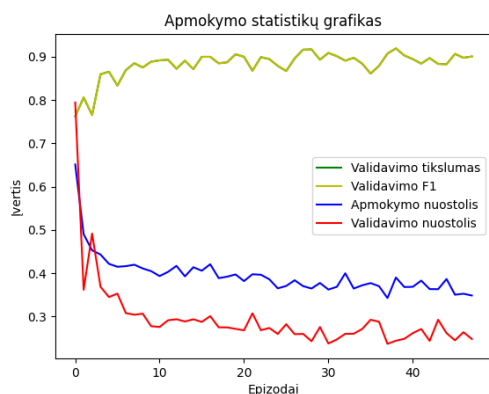
Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	83.83%	0.21191	81.56%	0.24351	73.09%	0.40393
M3	72.08%	0.41964	82.50%	0.24216	73.38%	0.43452
M4	81.04%	0.25269	82.56%	0.23222	72.77%	0.44329
M5	84.18%	0.22130	84.59%	0.21551	72.60%	0.36378

Apibendrinimas

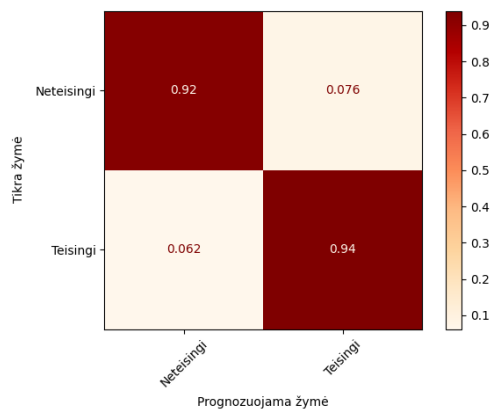
Iš suvidurkinto palyginimo, lentelė 5, galima pamatyti, kad vidutiniškai gerai pasirodė 3 modeliai M2, M3, M5. Vidutiniškai geriausias modelis yra M2 su **73.46%** tikslumu. Papildomos informacijos turėjimas apie pozos taškus pagerina modelio tikslumą vidutiniškai **1,31%**.

Palyginus tarpusavyje Tinetti pratimų rezultatus, matyti, kad didžiausias tikslumas pasiektas taikant modelį M5 Tinetti-9 pratimui. Tada Tinetti-1 ir Tinetti-2 klasifikavimo pratimų didžiausi tikslumo įverčiai pasiekti taikant M2 modelį, atitinkamai 75.11% ir 64.92% tikslumas. Tinetti-8 klasifikavimo didžiausias 72.62% tikslumas yra pasiektas su M3 modeliu. Tačiau verta atkreipti dėmesį, kad tai tik 0.35% geriau nei naudojant M2 modelį.

Paveikslėliuose 9 ir 10 yra pateikta Tinetti-9 M5 modelio apmokymo statistika bei klasifikavimo lentelė. Iš jų matosi, kad modelis pakankamai užtikrintai geba įvertinti Tinetti-9 pratimo atlikimo teisingumą.



9 pav. Tinetti-9 M5 modelio apmokymo ir validavimo statistikos



10 pav. Tinetti-9 M5 modelio klasifikavimo lentelė

5.8. Daugelio klasių klasifikavimo eksperimentai

5.8.1. Tikslas

Šių eksperimentų tikslas yra apmokyti modelį, kuris gebėtų atskirti atliekamus Tinetti-1, Tinetti-2, Tinetti-8, Tinetti-9 pratimus.

5.8.2. Konfigūracija

Pasirinkta optimizatoriaus funkcija - Adam. Pasirinkta nuostolių funkcija - kryžminė entropija. Mokama dinaminį skaičių epizodų stebint nuostolio funkcijos validavimą ir tikrinant su 10 epizodų uždelsimu (angl. patience) iki kada validavimo nuostolis mažėja. Apmokama pasirinkus $DR_s = [4, 8, 16, 32, 64, 128]$ duomenų rinkinio dydžius (angl. batch size).

5.8.3. Rezultatai

10 lentelė. Vidutiniškai 3-ų geriausių modelių rezultatai naudojant apmokymui duomenis be žmogaus pozos taškų matomumo, su matomumo įverčiais ir naudojant duomenų rinkinį su praplėtimu palyginimas

Modelio kodas	Be matomumo įverčių		Su matomumo įverčiais		Su praplėtimu ir matomumo įverčiais	
	F1	Validavimo nuostolis	F1	Validavimo nuostolis	F1	Validavimo nuostolis
M2	68.1%	0.83876	68.98%	0.7814	59.90%	1.1587
M3	68.46%	0.7136	69.69%	0.69162	61.50%	1.52038
M4	69.24%	0.76884	68.19%	0.76022	57.61%	1.27872
M5	68.53%	0.67852	70.78%	0.61424	62.01%	0.93777

Iš rezultatų lentelės 10 matoma, kad tarp modelių, apmokytų naudojant duomenų rinkinį be matomumo įverčio, vidutiniškai aukščiausią F1 įverti turi M4 modelis 68.24%, tačiau pastebėtina, kad nuo žemiausio modelio F1 įverčio skirtumas yra tik 1.14% ir skirtumai nėra reikšmingi. M5 modelio nuostolio funkcijos vidutiniškai nuostolis yra mažiausias tarp visų 0.67852, kas reikštų šio modelio patikimesnius rezultatus. Pridėjus matomumo įverčius ir pakartojus eksperimentus, bendra tendencija, išskyrus M4 modelį (sumažėjo 1.05%), yra vidutiniškai padidėjęs F1 įvertis, M2 ir M3 modelių atitinkamai 0.88% ir 1.23%, M5 modelį **2.25%**. Praplėtus Tinetti duomenų rinkinį papildomais duomenų įrašais iš praplėtimo ir atlikus eksperimentus naudojant pilną informaciją (matomumo įverčius), visų modelių F1 įvertis pastebimai sumažėjo, o nuostolio funkcijos įverčiai padidėjo. M2, M3 ir M4 F1 modelių vidutiniai įverčiai sumažėjo atitinkamai 9.08%, 8.19% ir 10.58%. Modelis M5 pasiekė aukščiausią vidutinį F1 įvertį **62.01%** palyginus su kitais modeliais.

5.9. Daugiatikslio klasifikavimo eksperimentai

5.9.1. Tikslas

Šių eksperimentų tikslas yra apmokyti modelį, kuris gebėtų atskirti atliekamus Tinetti-1, Tinetti-2, Tinetti-8, Tinetti-9 pratimus bei nustatyti ar pratimas yra atliekamas teisingai arba neteisingai.

5.9.2. Modeliai

Kiekviename Multi modelyje yra 2 modelio atšakos. Viena modelio atšaka sprendžia daugelio pratimų klasifikavimą, o kita atšaka sprendžia teisingumo klasifikavimo užduotį. Klasifikavimui naudojama kryžminės entropijos aktyvacijos funkcija.

5.9.2.1. Multi1

Multi1 modelis, lentelė 11, yra sudarytas M2 modelio, lentelė 4, pagrindu, pritaikius dviejų klasių ir daugelio klasių klasifikavimui. Modelis yra iš 2 konvoliucinių sluoksnių, 4 rinkinio normalizavimo, 2 sutelkimo imant maksimalią reikšmę sluoksnių, 1 ištiesinimo ir 3 tankinimo (angl. Dense) sluoksnių. Iš viso yra 2023878 (7.72 MB) parametrų, apmokamų - 2023174 (7.72 MB), neapmokamų - 704 (2.75 KB).

5.9.2.2. Multi2

Multi2 modelis, lentelė 25, yra sudarytas Multi1, lentelė 11, modelio pagrindu. Yra patobulintos modelio atšakos, į kiekvieną pridėjus po 2 rinkinio normalizavimo ir 2 tankinimo sluoksnius (32 ir 12 vienetų (angl. units) reikšmėmis daugelio klasių klasifikavimo atšakoje, ir 32 bei 8 vienetų reikšmėmis dviejų klasių klasifikavimo atšakoje). Detaliau žiūrėti Multi2 modelio lentelę 25. Iš viso yra 2032426 (7.75 MB) parametrų, apmokamų - 2031554 (7.75 MB), neapmokamų - 872 (3.41 KB).

5.9.2.3. Multi3

Multi3 modelis, lentelė 26, yra sudarytas Multi1, lentelė 11, modelio pagrindu. Yra pridėti 2 atsitiktinio praretinimo sluoksniai po sutelkimo imant maksimalią reikšmę sluoksnių ir po 1 atsitiktinio praretinimo sluoksnių į modelio atšakas po rinkinio normalizavimo sluoksnių. Iš viso yra 2023366 (7.72 MB) parametrų, apmokamų - 2022918 (7.72 MB), neapmokamų - 448 (1.75 KB).

5.9.2.4. Multi4

Multi4 modelis, lentelė 27, yra sudarytas M6, lentelė 23, modelio pagrindu. Kiekvienoje modelio atšakoje yra po 3 rinkinio normalizavimo ir 3 tankinimo sluoksnių. Iš viso yra 25386 (99.16 KB) parametrų, apmokamų - 24514 (95.76 KB), neapmokamų - 872 (3.41 KB).

5.9.2.5. Multi5

Multi5 modelis, lentelė 28, yra sudarytas Multi3, lentelė 26, pagrindu, pakeitus RELU aktyvacijos funkcijas ELU. Modelio parametrų skaičius atitinka Multi3 modelio parametrų skaičių.

11 lentelė. Multi1 modelis

Transformacija	Dimencija	Parametrai	Sujungta su	Aktyvacijos funk.
Skeletonas (Input)	(990, 1)	0		
Convolutional	(988, 32)	128	Skeletonas	RELU
Batch normalization	(988, 32)	128	Convolutional	
Max pooling	(494, 32)	0	Batch normalization	
Convolutional_1	(492, 64)	6208	Max pooling	RELU
Batch normalization_1	(492, 64)	256	Convolutional_1	
Max pooling_1	(246, 64)	0	Batch normalization_1	
Flatten	(15744)	0	Max pooling_1	
Dense	(128)	2015360	Flatten	RELU
Batch normalization_2	(128)	512	Dense	
Batch normalization_3	(128)	512	Dense	
Dense (class_type)	(4)	516	Batch normalization_2	Softmax
Dense (correctness)	(2)	258	Batch normalization_3	Softmax

5.9.3. Konfigūracija

Pasirinkta optimizatoriaus funkcija - Adam su mokymo žingsniu 10^{-3} . Yra sukonfigūruotas mokymo žingsnio planuotojas "ReduceLROnPlateau" su žingsniu 10^{-1} stebint 5 epizodų uždelsimu ir apribojus mokymo žingsnio mažinimą iki 10^{-6} . Pasirinktos nuostolių funkcijos - kryžminė entropija daugelio klasių užduočiai ir dviejų klasių užduočiai dviejų klasių kryžminės entropijos versija. Mokama dinaminį skaičių epizodų stebint nuostolio funkcijos validavimą ir tikrinant su 15 epizodų uždelsimu (angl. patience) iki kada validavimo nuostolis mažėja. Apmokama pasirinkus $DR_s = [4, 8, 16, 32, 64, 128]$ duomenų rinkinio dydžius (angl. batch size).

5.9.4. Rezultatai

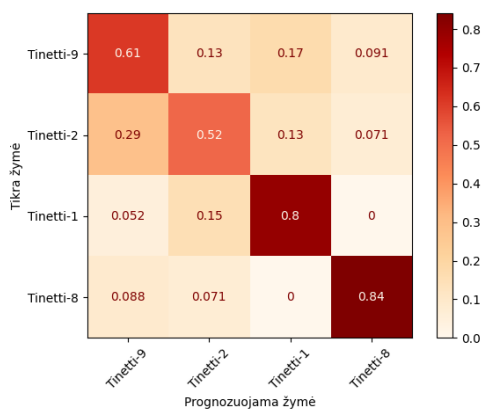
Eksperimentai buvo atlikti naudojant sukurtą Tinetti (skyrius 4.1) duomenų rinkinį ir pakartoti vėl prijungus duomenų rinkinio praplėtimą (skyrius 4.2). Atlikta 60 eksperimentų, po 30 su kiekvienu duomenų rinkiniu. Lentelėje 12 yra pateikiami suvidurkinti rezultatai pagal naudotą modelį kiekvienam duomenų rinkiniui.

12 lentelė. Suvidurkinti modelių rezultatai naudojant apmokymui Tinetti duomenų rinkinį bei naudojant Tinetti su praplėtimu duomenų rinkinį

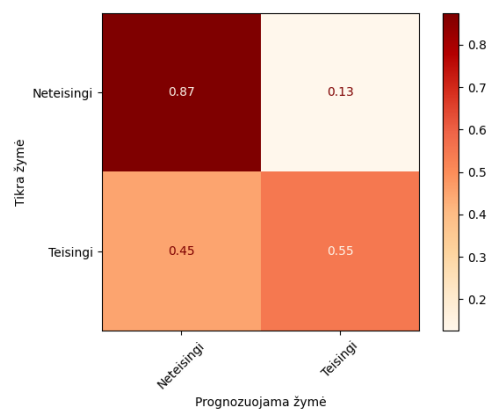
Modelio kodas	Tinetti duomenų rinkinys				Tinetti su praplėtimu			
	Pratimų F1	Teisingumo F1	Pratimų validavimo nuostolis	Teisingumo validavimo nuostolis	Pratimų F1	Teisingumo F1	Pratimų validavimo nuostolis	Teisingumo validavimo nuostolis
Multi1	70.85%	69.28%	0.70240	0.30093	61.97%	58.61%	1.29765	0.56857
Multi2	70.02%	70.40%	0.72786	0.2857	63.34%	56.97%	1.21561	0.4914
Multi3	70.77%	64.25%	0.60475	0.32045	59.38%	48.21%	0.94674	0.44629
Multi4	70.41%	62.78%	0.78	0.30443	63.07%	40.03%	2.01689	0.7752
Multi5	73.55%	63.21%	0.5846	0.32539	68.30%	45.34%	0.87924	0.42686

Iš rezultatų lentelėje 12 matosi, kad iki praplėtimo visi modeliai turi 1% ribose pratimų F1 70-71% tikslumą, išskyrus Multi5 modelį, kuris pasižymėjo 73,55% tikslumu. Geriausias 70.4% teisingumo F1 įvertis priklauso Multi2 modeliui. Po duomenų rinkinio praplėtimo Tinetti-1, Tinetti-2, Tinetti-9 vaizdo įrašais sumažėjo visų modelių tikslumas bei padidėjo nuostolis. Vidutiniškai aukščiausių pratimų F1 statistikos įvertį parodo Multi5 modelis, 68.30%, o vidutiniškai aukščiausių teisingumo F1 statistikos įvertį, 58.61%, turi Multi1 modelis.

Paveikslėliuose 11 ir 12 yra pavaizduotos Multi5 modelio pratimo ir teisingumo klasifikavimo lentelės, apmokymui naudojant Tinetti duomenų rinkinį su praplėtimu ir matomumo įverčiais taikant 128 duomenų rinkinio dydį.



11 pav. Multi5 modelio, apmokinto naudojant Tinetti su praplėtimu duomenų rinkinį, pratimų klasifikavimo lentelė (128 duomenų rinkinio dydis)



12 pav. Multi5 modelio, apmokinto naudojant Tinetti su praplėtimu duomenų rinkinį, teisingumo klasifikavimo lentelė (128 duomenų rinkinio dydis)

Lentelėje 13 pateikiamas Multi5 modelio, apmokyto naudojant Tinetti duomenų rinkinį su praplėtimu ir matomumo įverčiais, dviejų klasių klasifikavimo užduoties klasifikavimo raportas. Matoma, kad modelis "Neteisinga" klasei atlieka teisingų teigiamų prognozių daugiau negu "Teisinga" klasei, 88% prieš 54%. Jautrumas skiriasi panašiai, 87% prieš 55%, F1 statistika - 88% prieš 54%. Duomenų "Neteisinga" klasei yra apytiksliai 3,7 karto daugiau negu "Teisinga" klasei.

13 lentelė. Multi5 modelio, apmokyto Tinetti su praplėtimu duomenų rinkinio pagrindu, dviejų klasių klasifikavimo raportas

Klasė	Preciziškumas	Jautrumas	F1	# sk.
Neteisinga	0.88	0.87	0.88	2903
Teisinga	0.54	0.55	0.54	784
Tikslumas			0.81	3687
Makro Vid	0.71	0.71	0.71	3687
Pasvertas Vid	0.81	0.81	0.81	3687

Lentelėse 14 ir 15 pateikiamas Multi5 modelio, sprendžiančio daugelio klasių klasifikavimo užduotį, klasifikavimo raportas, atitinkamai, apmokytas naudojant Tinetti be praplėtimo ir Tinetti su praplėtimu duomenų rinkinius su duomenų rinkinio dydžiu 128. Matoma, kad po praplėtimo Tinetti-2, Tinetti-8 ir Tinetti-9 vaizdo įrašais, modelio F1 tikslumas bei vidurkiai sumažėjo 0,01-0,02 ribose.

14 lentelė. Multi5 modelio, apmokyto naudojant Tinetti be praplėtimo duomenų rinkinį, pasirinkus duomenų rinkinio 128 dydį, daugelio klasių klasifikavimo raportas

Klasė	Preciziškumas	Jautrumas	F1	# sk.
Tinetti-9	0.49	0.86	0.63	838
Tinetti-2	0.73	0.49	0.58	1033
Tinetti-1	0.65	0.35	0.46	342
Tinetti-8	0.96	0.88	0.92	1221
Tikslumas			0.70	3434
Makro Vid	0.71	0.64	0.65	3434
Pasvertas Vid	0.75	0.70	0.70	3434

15 lentelė. Multi5 modelio, apmokyto naudojant Tinetti su praplėtimu duomenų rinkinį, pasirinkus duomenų rinkinio 128 dydį, daugelio klasių klasifikavimo raportas

Klasė	Preciziškumas	Jautrumas	F1	# sk.
Tinetti-9	0.56	0.61	0.58	934
Tinetti-2	0.69	0.52	0.59	1132
Tinetti-1	0.51	0.80	0.62	400
Tinetti-8	0.86	0.84	0.85	1221
Tikslumas			0.68	3687
Makro Vid	0.65	0.69	0.66	3687
Pasvertas Vid	0.69	0.68	0.68	3687

5.10. Agreguotų kadrų eksperimentai

Šiuose eksperimentuose siekiama pamatyti kaip kinta modelio prognozės, duomenų paruošimo žingsnyje taikant 10 ir 25 iš eilės einančių kadrų transformacijas. Transformacijos žingsnyje yra formuojama apmokymui skirta eilutė kaip parodyta formulėje 22.

Apmokamas Multi5 modelis su $DR_s = [4, 8, 16, 32, 64, 128]$ duomenų rinkinio dydžiais, naudojant Tinetti duomenų rinkinį be praplėtimo su matomumo įverčiais. Lentelėje 16 yra pateikiami pratimų F1 ir teisingumo F1 metrikų palyginimo rezultatai.

16 lentelė. Multi5 modelio metrikų palyginimas su pritaikytomis 10 ar 25 kadrų agregavimo transformacijomis

Duomenų rinkinio dydis	Agreguotų kadrų skaičius	Pratimų F1	Teisingumo F1	Pratimų nuostolis	Teisingumo nuostolis
4	10	71.28%	57.47%	0.63977	0.35394
4	25	73.92%	59.14%	0.59200	0.35570
8	10	72.33%	59.20%	0.58678	0.33627
8	25	77.52%	61.95%	0.49800	0.32893
16	10	75.01%	63.67%	0.56423	0.32172
16	25	77.65%	65.70%	0.49957	0.32476
32	10	74.67%	64.43%	0.56138	0.31384
32	25	74.49%	62.68%	0.54750	0.33650
64	10	75.02%	65.74%	0.58158	0.31261
64	25	49.78%	46.49%	1.27282	0.73865
128	10	75.30%	65.61%	0.57009	0.30965
128	25	43.58%	47.62%	1.76053	0.71206

Iš lentelės 16 matosi, kad 10 agreguotų kadrų atveju, didinant duomenų rinkinio dydį, pratimų F1 statistika nuosekliai didėja (išskyrus 16 duomenų rinkinio dydžio atvejį), kol pasiekia aukščiausią įvertį 75.30%. Teisingumo F1 statistika taip pat nuosekliai didėja nuo 4 iki 64 duomenų rinkinio dydžio kai pasiekia 65.74%, o vėliau pradeda mažėti. 25 agreguotų kadrų atveju, aukštesni pratimų F1 statistikos įverčiai yra gaunami taikant 4, 8, 16, 32 duomenų rinkinio dydžius (aukščiausias pasiektas su 16 duomenų rinkinio dydžiu - 77.65%), o teisingumo F1 statistikos įverčiai nuosekliai didėja iki duomenų rinkinio dydžio 16 kai pasiekia aukščiausią įvertį 65.70%, o toliau pradeda mažėti.

Apibendrinant lentelės 16 rezultatus, aukščiausią pratimų F1 įvertį pavyko pasiekti pritaikius 25 kadrų transformaciją su duomenų rinkinio dydžiu 16 (pratimų 77.65%, teisingumo - 65.7%). Aukščiausi teisingumo F1 įverčiai gaunami, pritaikius 10 kadrų transformaciją su 64 duomenų rinkinio dydžiu (pratimų - 75.02%, teisingumo - 65.74%), ir pritaikius 25 kadrų transformaciją su 16 duomenų rinkinio dydžiu (pratimų - 77.65%, teisingumo - 65.70%).

5.11. Neteisingų atvejų testavimas

Neteisingų atvejų testavimui yra naudojami vaizdo įrašai iš Tinetti duomenų rinkinio praplėtimo bei keli vaizdo įrašai iš duomenų rinkinio testavimo pogrupio.

Validavimas atliktas naudojant Multi5 modelį, kuris apmokytas naudojant Tinetti duomenų rinkinį be praplėtimo ir su matomumo įverčiais, su pritaikyta 10 kadro agregavimo transformacija ir pritaikius 128 duomenų rinkinio dydį. Rezultatai pateikti lentelėje 17.

17 lentelė. Multi5 modelio prognozės ant validavimo vaizdo įrašų

Vaizdo įrašas	Atitikimo Tinetti pratimams	Etalono korektiškumas	Pratimų prognozės	Korektiškumo prognozės
a16_s03_e02_rgb.avi	Tinetti 9	Neteisingas	[3 3 3 3 3 3 3 3 3 1 1 2 2 1 2 2]	89 - 99% neteisinga
a16_s06_e02_rgb.avi	Tinetti 9	Neteisingas	[3 3 3 3 3 3 3 3]	89 - 99% neteisinga
a16_s01_e01_rgb.avi	Tinetti 9	Neteisingas	[3 3 3 3 2 1 0 0 0 0 0 0 0]	89 - 99% neteisinga
a16_s08_e02_rgb.avi	Tinetti 9	Neteisingas	[3 3 3 0 1 1 2 2]	89 - 99% neteisinga
GestureRecording_Id13actor3idAction13category6.avi	Tinetti 9	Neteisingas	[0 0]	[0.93 0.07], [0.85 0.15]
GestureRecording_Id16actor8idAction16category6.avi	Tinetti 9	Neteisingas	[0 1]	[1. 0.], [1. 0.]

Pastaba 1. Pastebėta, kad kai vaizdo įrašas prasideda nuo stovinčio žmogaus pozos ir kol dar veiksmas nėra prasidėjęs, modelis priskiria kadro grupę Tinetti-8 pratimui bei nustato jų atlikimo korektiškumą kaip neteisingą.

Pastaba 2. Pastebėta, kad modelis kai kada skirtingas vaizdo įrašo dalis priskiria skirtingiems pratimams bei nustato kitokį jo korektiškumą. Pavyzdžiui, vaizdo įrašo "a16_s10_e01_rgb.avi" kadro sekoms (kurių vieną sudaro 10 nuoseklių kadro) pratimų prognozės [3 3 3 3 0 2 2 2 2] su atitinkamomis neteisingumo prognozėmis [0.9, 0.91, 0.87, 0.67, 0.01, 0.07, 0.43, 0.1, 0.1]. Iš jų matyti, kad vaizdo įrašo pradžioje judesys neaptiktas (remiantis pirma pastaba, pratimo klasės prognozė - 3 ir neteisingumo 90% įvertis), vėliau seka atsistėjimo veiksmas (klasė - 2) įvertintas kaip teisingai atliekamas su tikimybe 0.99%, bei vaizdo įrašas užsibaigia pusiausvyros laikymu sėdint, kas atitinka klasę 2 ir įvertinta vidutiniškai teisingai.

Pastaba 3. Pastebėta, kad kartais pasitaiko tokių kadro sekų, kur neaišku ar judesys priklauso atsistėjimo ar atsistojimo veiksmui.

Toliau pavaizduota lentelė 18, kurioje yra palyginami gautos modelio prognozės prieš ir po atliktų pasirinkto kūno pozos taškų koregavimų.

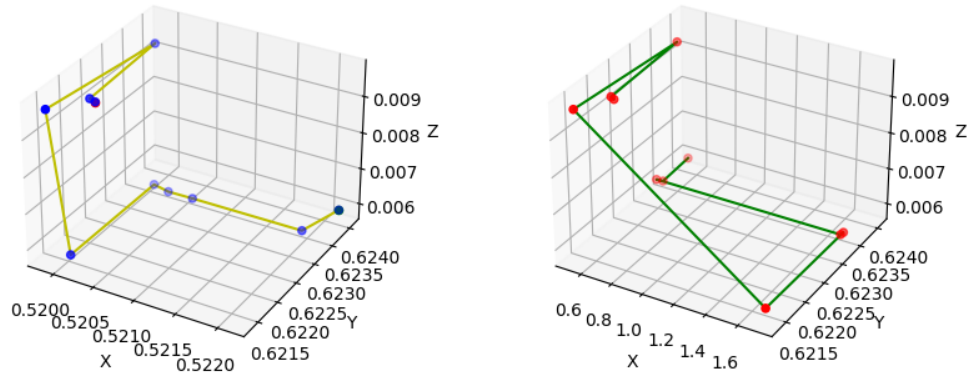
18 lentelė. Modelio prognozių palyginimai prieš ir po pasirinktų kūno taškų koordinatinių modifikavimų

Vaizdo įrašas	Keičiama kūno pozos dalis	Iki modifikavimo		Po modifikavimo	
		Pratimų prognozė	Korektiškumo prognozė	Pratimų prognozė	Korektiškumo prognozė
a16_s08_e02_rgb.avi	Kairysis petis	Tinetti-1	[0.45 0.55]	Tinetti-1	[0. 1.]
a16_s08_e02_rgb.avi	Dešinysis kelias	Tinetti-1	[0.45 0.55]	Tinetti-1	[0.61 0.39]
GestureRecording_Id15actor3idAction15category7.avi	Kairysis klubas	Tinetti-2	[0.99 0.01]	Tinetti-2	[0.33 0.67]

Iš lentelės 18 rezultatų matoma, kad atlikus kairiojo peties, dešiniojo kelio ir kairiojo klubo x koordinatės pokyčius, modelio teisingumo įvertis pasikeitė. Vaizdo įrašui *a16_s08_e02_rgb.avi*

kairiojo peties atveju iš tikimybės 55% į 100 %, kad teisingai atliktas. Dešiniojo kelio postūmiui x-ašyje padidėjo tikimybė, kad pratimas neteisingai atliktas iš 45% iki 61%. Kairiojo klubo padėties keitimas, pav. 13, įtakojo teisingo atlikimo tikimybės padidėjimą iki 67%.

Kairiojo klubo x koordinatės pokytis erdvėje

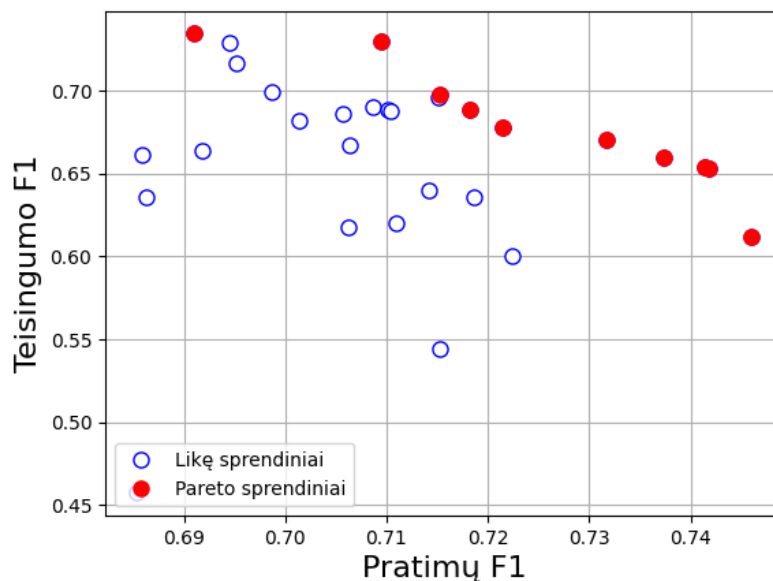


13 pav. Validavimo vaizdo įrašo žmogaus kairiojo klubo taško koordinatinių pokyčių demonstravimas

5.12. Daugiatikslio klasifikavimo optimalių sprendinių paieška

Daugiatikslio klasifikavimo Pareto optimalių sprendinių paieška vykdyta ieškant optimalių pratimų ir pratimų teisingumo F1 statistikų reikšmių kombinacijų.

Pareto optimalūs sprendiniai pateikti lentelėje 19 modeliams, apmokytiems Tinetti be praplėtimo su matomumo įverčiais duomenų rinkinio pagrindu. Pav. 14 pateiktas optimalių sprendinių projekcija.



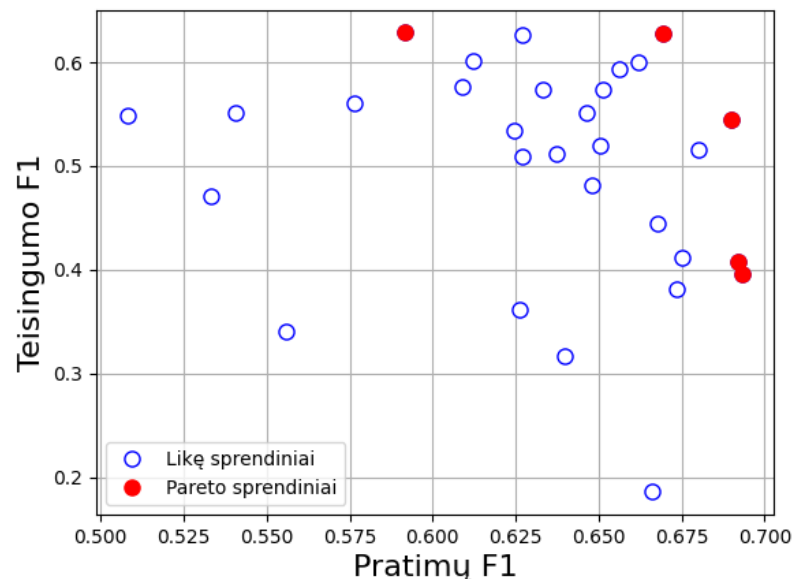
14 pav. Pareto optimalių sprendinių projekcija modeliams, apmokytiems naudojant Tinetti su matomumo įverčiais duomenų rinkinį be praplėtimo

19 lentelė. Pareto optimalūs sprendiniai modeliams, apmokytiems Tinetti su matomumo įverčiais be praplėtimo duomenų rinkinio pagrindu

Modelis	Duomenų rinkinio dydis	Pratimų F1	Teisingumo F1	Pratimų nuostolis	Teisingumo nuostolis
Multi2	64	69.11%	73.43%	0.82335	0.26093
Multi1	64	70.94%	72.95%	0.76358	0.30618
Multi1	16	71.53%	69.76%	0.67297	0.29297
Multi1	8	71.82%	68.88%	0.63161	0.27754
Multi4	16	72.15%	67.8%	0.70986	0.26822
Multi5	128	73.17%	67.02%	0.58073	0.31111
Multi5	64	73.74%	65.96%	0.57435	0.31156
Multi5	16	74.14%	65.35%	0.57583	0.32234
Multi5	32	74.18%	65.35%	0.56681	0.31952
Multi5	8	74.6%	61.16%	0.58265	0.33425

Iš lentelės 19 rezultatų matoma, kad didžiausias pratimų F1, 74.6%, įvertis pasiektas naudojant Multi5 modelį su 8 duomenų rinkinio dydžiu. Didžiausias teisingumo F1, 73.43%, įvertis pasiektas su Multi2 modeliu naudojant 64 duomenų rinkinio dydį.

Pareto optimalūs sprendiniai pateikti lentelėje 20 modeliams, apmokytiems Tinetti su matomumo įverčiais ir praplėtimu. Matoma pav. 15, kad buvo rasti 5 optimalūs sprendiniai.



15 pav. Pareto optimalių sprendinių projekcija modeliams, apmokytiems naudojant Tinetti su matomumo įverčiais ir praplėtimu duomenų rinkinį

20 lentelė. Pareto optimalūs sprendiniai modeliams, apmokytiems Tinetti su matomumo įverčiais ir praplėtimu duomenų rinkinio pagrindu

Modelis	Duomenų rinkinio dydis	Pratimų F1	Teisingumo F1	Pratimų nuostolis	Teisingumo nuostolis
Multi2	4	66.95%	62.81%	0.86006	0.37151
Multi1	32	59.16%	62.85%	1.18169	0.51319
Multi5	32	69.34%	39.54%	0.90379	0.42507
Multi5	64	69.2%	40.76%	0.90322	0.41451
Multi5	128	68.98%	54.5%	0.92132	0.40832

Iš lentelės 20 matoma, kad didžiausias pratimų F1, 69.34%, yra pasiektas su Multi5 modeliu ir 32 duomenų rinkinio dydžiu, bet turi mažiausią teisingumo F1, 39.54%, įvertį. Didžiausias teisingumas F1, 62.85%, yra pasiektas, naudojant Multi1 modelį su 32 duomenų rinkinio dydžiu. Vidutiniškai gerai pasirodė Multi2 ir Multi5 modeliai, atitinkamai, pritaikius 4 ir 128 duomenų rinkinio dydžius.

Lyginant lentelių 19 ir 20 rezultatus, matoma, kad daugiatikslio uždavinio modeliai, apmokyti su duomenų rinkinio matomumo įverčiais ir praplėtimu, bendrai rodo mažesnius pratimo F1 ir teisingumo F1 įverčius už apmokytus be praplėtimu.

Rezultatai ir išvados

Rezultatai

1. Sukurtas naujas Tinetti duomenų rinkinys į kurį įeina Tinetti-1 (98 vaizdo įrašų), Tinetti-2 (177 vaizdo įrašų), Tinetti-8 (137 vaizdo įrašų), Tinetti-9 (169 vaizdo įrašų) pratimai. Iš viso 581 vaizdo įrašas.
2. Praplėstas naujai sukurtas Tinetti duomenų rinkinys papildomais įrašais iš "Florence 3D Action" - 26 vaizdo įrašų ir "MSR DailyActivity3D" - 46 vaizdo įrašų. Iš viso 72 vaizdo įrašai.
3. Pagrindiniai identifikuoti rezultatai
 - 3.1. naudojant sukurtą Tinetti duomenų rinkinį be matomumo įverčių ir su matomumo įverčiais parodė:
 - 3.1.1. Dviejų klasių atveju. Tinetti-1, Tinetti-2, Tinetti-8 pratimams M2 modelio F1 statistikos 2.21%, 4.05% ir 1.16% pagerėjimą. Tinetti-8 ir Tinetti-9 pratimams M5 modelio F1 statistikos 3.39% ir 0.41% pagerėjimą.
 - 3.1.2. Daugelio klasių atveju M5 modelio F1 statistikos gerėjimą 2.25%.
 - 3.2. naudojant Tinetti duomenų rinkinį su matomumo įverčiais ir naudojant Tinetti duomenų rinkinį su matomumo įverčiais ir praplėtimu parodė:
 - 3.2.1. Dviejų klasių atveju. Tinetti-1 pratimui M2 modelio - 6,66%, o M5 modelio - 13,89% F1 statistikos pagerėjimus.
4. Pasiūlytos modelių variacijos (Multi) skirtos identifikuoti atliekamą Tinetti pratimą ir jo korektiškumą.
 - 4.1. Su Multi5 modeliu pasiekta pratimų F1 statistikos 73.55% įvertį, o su Multi2 teisingumo F1 statistikos 70.40% įvertį suvidurkintuose rezultatuose, naudojant Tinetti duomenų rinkinį su matymo įverčiais be praplėtimo.
5. Multi konfigūracijos atveju rasti optimalūs Pareto Fronto sprendiniai.
6. Atlikti eksperimentai su Multi5 modeliu apmokymui naudojant Tinetti duomenų rinkinį su matomumo įverčiais be praplėtimo ir pakeičiant agreguotų kadrų kiekį į 25 duomenų paruošimo žingsnyje parodė, kad aukščiausi modelio pasiekti pratimo F1 - 77.65% ir korektiškumo F1 - 65.70% statistikų įverčiai yra gaunami, naudojant 16 duomenų rinkinio dydį. Pritaikius 10 kadrų agregavimo transformaciją, aukščiausi modelio pratimo F1 - 75.30% ir korektiškumo - 65.74% statistikų įverčiai yra pasiekti, naudojant, atitinkamai, 128 ir 64 duomenų rinkinio dydžius.

Išvados

1. Duomenų rinkinio papildymas su MediaPipe BlazePose apskaičiuotais žmogaus pozos taškų matomumo įverčiais modelio klasifikavimo rezultatus pagerina.
2. Sukurtas daugiatis modelis įgalina spręsti du klasifikavimo uždavinius vienu metu taip apibendrinant modelių veikimą.
3. Atlikus Pareto Fronto optimizavimo analizę, identifikuota 10 praktiškų modelių, apmokytų naudojant Tinetti duomenų rinkinį su matomumo įverčiais ir 5 modeliai, apmokyti naudojant Tinetti duomenų rinkinį su matomumo įverčiais ir praplėtimu.
4. Praktiškai geresni klasifikavimo rezultatai gaunami taikant 25 kadrų agregavimo transformaciją duomenų paruošimo žingsnyje ir taikant 16 duomenų rinkinio dydį, naudojant Tinetti duomenų rinkinį su matomumo įverčiais be praplėtimo.

Šaltiniai

- [ACN⁺20] M. Al-Faris, J. Chiverton, D. Ndzi, A. I. Ahmed. A review on computer vision-based methods for human action recognition. *Journal of imaging*. 2020, tomas 6, numeris 6, p. 46.
- [AJL⁺21] T. Ahmad, L. Jin, L. Lin, G. Tang. Skeleton-based action recognition using sparse spatio-temporal GCN with edge effective resistance. *Neurocomputing*. 2021, tomas 423, p. 389–398.
- [AKS⁺21] F. Afza, M. A. Khan, M. Sharif, S. Kadry, G. Manogaran, T. Saba, I. Ashraf, R. Damaševičius. A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. *Image and Vision Computing*. 2021, tomas 106, p. 104090.
- [AR11] J. K. Aggarwal, M. S. Ryoo. Human activity analysis: A review. *Acm Computing Surveys (Csur)*. 2011, tomas 43, numeris 3, p. 1–43.
- [Arm16] B. Armacka. *Greitėjančio tempo treniruotės poveikis pacientų po galvos smegenų insulto pusiausvyrai ir eisenai*. 2016. Magistro darbas. Vilniaus universitetas.
- [BB08] S. Barris, C. Button. A review of vision-based motion analysis in sport. *Sports medicine*. 2008, tomas 38, p. 1025–1043.
- [BGR⁺20] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, M. Grundmann. BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*. 2020.
- [BMK⁺17] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, R. Jenssen. Recurrent neural networks for short-term load forecasting: an overview and comparative analysis. 2017.
- [CCL⁺09] C.-W. Cho, W.-H. Chao, S.-H. Lin, Y.-Y. Chen. A vision-based analysis system for gait recognition in patients with Parkinson's disease. *Expert Systems with applications*. 2009, tomas 36, numeris 3, p. 7033–7039.
- [CEC⁺18] S. L. Colyer, M. Evans, D. P. Cosker, A. I. Salo. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports medicine-open*. 2018, tomas 4, p. 1–15.
- [C]JK15] C. Chen, R. Jafari, N. Kehtarnavaz. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. Iš: *2015 IEEE International conference on image processing (ICIP)*. IEEE, 2015, p. 168–172.
- [CSW⁺17] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. Iš: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, p. 7291–7299.

- [CUH15] D.-A. Clevert, T. Unterthiner, S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*. 2015.
- [Cui18] N. Cui. Applying gradient descent in convolutional neural networks. Iš: *Journal of Physics: Conference Series*. IOP Publishing, 2018, tomas 1004, p. 012027.
- [CVG⁺14] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. 2014.
- [DBZ23] Q. Du, H. Bai, Z. Zhu. Intelligent Evaluation Method of Human Cervical Vertebra Rehabilitation Based on Computer Vision. *Sensors*. 2023, tomas 23, numeris 8, p. 3825.
- [FXT⁺17] H.-S. Fang, S. Xie, Y.-W. Tai, C. Lu. Rmpe: Regional multi-person pose estimation. Iš: *Proceedings of the IEEE international conference on computer vision*. 2017, p. 2334–2343.
- [GBC16] I. Goodfellow, Y. Bengio, A. Courville. *Deep learning*. MIT press, 2016.
- [HG16] D. Hendrycks, K. Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. 2016.
- [HS97] S. Hochreiter, J. Schmidhuber. Long short-term memory. *Neural computation*. 1997, tomas 9, numeris 8, p. 1735–1780.
- [HZR⁺15] K. He, X. Zhang, S. Ren, J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Iš: *Proceedings of the IEEE international conference on computer vision*. 2015, p. 1026–1034.
- [YND⁺18] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*. 2018, tomas 9, p. 611–629.
- [Isa23] Z. Isaac. *Overview of Rehabilitation* [<https://www.merckmanuals.com/professional/special-subjects/rehabilitation/overview-of-rehabilitation/>]. 2023.
- [YSG⁺19] S. K. Yadav, A. Singh, A. Gupta, J. L. Raheja. Real-time Yoga recognition using deep learning. *Neural computing and applications*. 2019, tomas 31, p. 9349–9361.
- [JKA⁺20] I. Jegham, A. B. Khalifa, I. Alouani, M. A. Mahjoub. Vision-based human action recognition: An overview and real world challenges. *Forensic Science International: Digital Investigation*. 2020, tomas 32, p. 200901.
- [JNB19] A. Jalal, A. Nadeem, S. Bobasu. Human body parts estimation and detection for physical sports movements. Iš: *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*. IEEE, 2019, p. 104–109.

- [Jun18] M. Junka. *Virtualios realybės priemonių taikymas reabilitacijoje pacientams po galvos smegenų insulto. Sisteminė literatūros apžvalga*. 2018. Disertacija. Lietuvos sveikatos mokslų universitetas.
- [KMK⁺16] D. Krueger, T. Maharaj, J. Kramár, M. Pezeshki ir kiti. Zoneout: Regularizing rnns by randomly preserving hidden activations. *arXiv preprint arXiv:1606.01305*. 2016.
- [Lan13] A. Lančinskas. Atsitiktinės paieškos globaliojo optimizavimo algoritmų lygiagretinimas. 2013.
- [LBB⁺98] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998, tomas 86, numeris 11, p. 2278–2324.
- [LDK⁺20] C.-B. Lin, Z. Dong, W.-K. Kuan, Y.-F. Huang. A framework for fall detection based on OpenPose skeleton and LSTM/GRU models. *Applied Sciences*. 2020, tomas 11, numeris 1, p. 329.
- [LHW⁺21] X. Li, Y. Hou, P. Wang, Z. Gao, M. Xu, W. Li. Trear: Transformer-based rgb-d egocentric action recognition. *IEEE Transactions on Cognitive and Developmental Systems*. 2021, tomas 14, numeris 1, p. 246–252.
- [LSP⁺19] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, A. C. Kot. Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. *IEEE transactions on pattern analysis and machine intelligence*. 2019, tomas 42, numeris 10, p. 2684–2701.
- [LSS⁺19] L. Lu, Y. Shin, Y. Su, G. E. Karniadakis. Dying relu and initialization: Theory and numerical examples. *arXiv preprint arXiv:1903.06733*. 2019.
- [LWW⁺19] Y. Liu, X. Wang, L. Wang, D. Liu. A modified leaky ReLU scheme (MLRS) for topology optimization with multiple materials. *Applied Mathematics and Computation*. 2019, tomas 352, p. 188–204.
- [Nie15] M. A. Nielsen. *Neural networks and deep learning*. Tomas 25. Determination press San Francisco, CA, USA, 2015.
- [PBK⁺12] K. Pulli, A. Baksheev, K. Korniyakov, V. Eruhimov. Real-time computer vision with OpenCV. *Communications of the ACM*. 2012, tomas 55, numeris 6, p. 61–69.
- [QKG19] Z. Qin, D. Kim, T. Gedeon. Rethinking softmax with cross-entropy: Neural network classifier as mutual information estimator. *arXiv preprint arXiv:1911.10688*. 2019.
- [QSS⁺18] D. Qurratu'aini, A. Sophian, W. Sediono, H. M. Yusof, S. Sudirman. Visual-based fingertip detection for hand rehabilitation. *Indones. J. Electr. Eng. Comput. Sci.* 2018, tomas 9, numeris 2, p. 474.
- [RAM16] S. Ranasinghe, F. Al Machot, H. C. Mayr. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks*. 2016, tomas 12, numeris 8, p. 1550147716665520.

- [RHW86] D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning representations by back-propagating errors. *nature*. 1986, tomas 323, numeris 6088, p. 533–536.
- [RSM⁺18] M. A. F. Rodrigues, Y. R. Serpa, D. V. Macedo, E. S. Sousa. A serious game to practice stretches and exercises for a correct and healthy posture. *Entertainment Computing*. 2018, tomas 28, p. 78–88.
- [RZL17] P. Ramachandran, B. Zoph, Q. V. Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*. 2017.
- [SHK⁺14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*. 2014, tomas 15, numeris 1, p. 1929–1958.
- [SLC04] C. Schuldt, I. Laptev, B. Caputo. Recognizing human actions: a local SVM approach. Iš: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. IEEE, 2004, tomas 3, p. 32–36.
- [SLN⁺16] A. Shahroudy, J. Liu, T.-T. Ng, G. Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. Iš: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 1010–1019.
- [SP21] A. Stergiou, R. Poppe. Learn to cycle: Time-consistent feature discovery for action recognition. *Pattern Recognition Letters*. 2021, tomas 141, p. 1–7.
- [STA12] G. Sebestyen, A. Tirea, R. Albert. Monitoring Human Activity through Portable Devices. *Carpathian Journal of Electronic & Computer Engineering*. 2012, tomas 5, numeris 1.
- [SVB⁺13] L. Seidenari, V. Varano, S. Berretti, A. Bimbo, P. Pala. Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. Iš: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2013, p. 479–485.
- [VDD15] K. Vamsikrishna, D. P. Dogra, M. S. Desarkar. Computer-vision-assisted palm rehabilitation with supervised learning. *IEEE Transactions on Biomedical Engineering*. 2015, tomas 63, numeris 5, p. 991–1001.
- [WG15] H. Wu, X. Gu. Towards dropout training for convolutional neural networks. *Neural Networks*. 2015, tomas 71, p. 1–10.
- [WYX⁺21] H. Wang, B. Yu, K. Xia, J. Li, X. Zuo. Skeleton edge motion networks for human action recognition. *Neurocomputing*. 2021, tomas 423, p. 1–12.
- [WLW⁺12] J. Wang, Z. Liu, Y. Wu, J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. Iš: *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, p. 1290–1297.

Priedai

Priedas nr. 1

TINETTI Testas

Pusiausvyros vertinimas pagal Tinetti skalę		
Pradinė padėtis – žmogus sėdi ant kietos kėdės, kėdė be laikiklių rankoms.		
Nr.	Pavadinimas	Vertinimas (balais)
1	Pusiausvyros laikymas sėdint	- svyruoja į šonus ar slidinėja kėdėje - 0 - sėdi saugiai, stabiliai - 1
2	Atsistojimas iš sėdimos padėties	- reikalinga pagalba - 0 - atsistoja pats, tačiau su rankų pagalba - 1 - savarankiškai be rankų pagalbos - 2
3	Bandymas atsistoti	- reikalinga pagalba - 0 - savarankiškai, tačiau ne iš pirmo karto - 1 - savarankiškai, iš pirmo karto - 2
4	Pusiausvyros laikymas pirmąsias 5s po atsistojimo	- nestabiliai (svyruoja į šalis, stato koją į šalį) - 0 - stabiliai, bet naudoja vaikštynę ar lazda - 1 - stabiliai, be pagalbinių priemonių - 2
5	Pusiausvyros laikymas stovint	- nestabiliai - 0 - stabiliai, tačiau reikalinga plati atraminė plokštuma (kojos pastatytos plačiai) arba reikia lazdos ar kitų pagalbinių priemonių - 1 - siaura atraminė plokštuma, be pagalbinių priemonių - 2
6	Stumtelėjimas (stovima suglaustomis pėdomis, tyrėjas lengvai stumteli delnu tiriamąjį krūtinkaulio srityje 3 kartus)	- krenta - 0 - svyruoja, ieško atramos - 1 - stovi stabiliai - 2
7	Stumtelėjimas (stovima kaip Nr.6, tik užmerktomis akimis)	- stovi nestabiliai - 0 - stovi stabiliai - 1
8	Apsisukimas 360° kampu	- nutrūkstančiais žingsniais (t.y. žengia porą žingsnių, sustoja, toliau tęsia) - 0 - nenutrūkstančiai žingsniuoja - 1
		- nestabiliai atlieka (svyruoja, ieško atramos) - 0 - atlieka stabiliai - 1
9	Atsisėdimas	- nesaugiai (neapskaičiuoja atstumo, krenta į kėdę) - 0 - rankų pagalba arba pats judesys nėra tolygus - 1 - atlieka saugiai, tolygiai - 2
Viso:		

Eisenos vertinimas pagal Tinetti skalę		
Atlikimo technika – einama skersai kambario, pirmyn įprastiniu žingsniu, grįžtant atgal greitai, bet saugiu žingsniu.		
10	Ėjimo pradžia (tuoj pat, kai pasakome žmogui, kad pradėtų eiti)	- dvejojimas, daugybiniai bandymai pradėti - 0 - be dvejonių, iš karto - 1
11	Žingsnio ilgis ir aukštis a) dešinės pusės mostas:	- darant mostą, pėda nepastatoma į priekį nuo kairės pėdos - 0 - mosto metu dešinė pėda yra priekyje kairės pėdos atžvilgiu - 1 - einant, dešinė koja pilnai neatkeliama nuo grindų - 0 - dešinė koja atkeliama - 1
	b) kairės kojos mostas:	- kairė pėda nėra pastatoma į priekį dešinės atžvilgiu - 0 - kairė pėda priekyje dešinės - 1 - kairė pėda neatkeliama - 0 - kairė pėda nevelkama grindimis - 1
12	Žingsnių simetriškumas	- dešinys ir kairys žingsniai nesimetriški - 0 - žingsniai vienodo ilgio - 1
13	Žingsniavimo tolygumas	- eina – sustoja, po to vėl eina, žingsniai nutrūkstantys - 0 - nenutrūkstamas žingsniavimas - 1
14	Ėjimas viena linija	- ryškus nukrypimas - 0 - nežymus nukrypimas arba naudojimas pagalbinėmis priemonėmis - 1 - ėjimas tiesiai be pagalbinių priemonių - 2
15	Kūnas	- ryškus svyravimas einant ar naudojimas pagalbinių priemonių - 0 - nėra svyravimo, bet einama sulenktais keliais, ištiestomis į šonus rankomis - 1 - nėra svyravimo, nereikalinga rankų pagalba, nereikia pagalbinių priemonių - 2
16	Pėdų padėtis	- einant kulnai toli vienas nuo kito - 0 - kulnai beveik liečia vienas kitą - 1
Viso:		
Bendra viso įvertinimo suma:		
<19 – yra didelė rizika nukristi 19-24 – yra nedidelė rizika nukristi 28 – norma		

Priedas nr. 2

Dviejų/daugelio klasių klasifikavimo modeliai

22 lentelė. M5 modelis

Transformacija	Dimencija	Filtrų sk.	Filtro dydis	Aktyvacijos funk.	Parametrai
Input	(990, 1)				
Convolutional	(988, 32)	32	3x3	ReLU	128
Batch normalization	(988, 32)	-	-	-	128
Max pooling	(494, 32)	-	2x2	-	0
Dropout	(494, 32)	-	-	-	0
Convolutional	(492, 64)	64	3x3	ReLU	6208
Batch normalization	(492, 64)	-	-	-	256
Max pooling	(246, 64)	-	2x2	-	0
Dropout	(246, 64)	-	-	-	0
Flatten	(15744)	-	-	-	0
Dense	(128)	-	-	ReLU	2015360
Batch normalization	(128)	-	-	-	512
Dropout	(128)	-	-	-	0
Dense	(2)	-	-	Softmax	258

23 lentelė. M6 modelis

Transformacija	Dimencija	Filtrų skaičius	Filtro dydis	Žingsnis	Aktyvacijos funk.	Parametrai
Input	(990, 1)					
Convolution	(988, 32)	32	3x3	1	Relu	128
Batch Normalization	(988, 32)	-	-	-	-	128
Max pooling	(494, 32)	-	-	2	-	0
Convolution	(492, 64)	64	3x3	1	Relu	6208
Batch Normalization	(492, 64)	-	-	-	-	256
Max pooling	(246, 64)	-	-	2	-	0
Global average pooling	(64)	-	-	-	-	0
Dense	(128)	-	-	-	Relu	8320
Batch Normalization	(128)	-	-	-	-	512
Dense	(2)	-	-	-	Softmax	258

24 lentelė. M7 modelis

Transformacija	Dimencija	Filtrų sk.	Filtrų dydis	Žingsnis	Aktyvacijos funk.	Parametrai
Input	(990, 1)	-	-	-	-	0
Convolutional	(988, 32)	32	3x3	1	Relu	128
Max pooling	(494, 32)	-	-	-	-	0
Convolutional	(492, 64)	64	3x3	1	Relu	6208
Max pooling	(246, 64)	-	-	-	-	0
Reshape	(246, 64)	-	-	-	-	0
Layer normalization	(246, 64)	-	-	-	-	128
Multi head attention	(246, 64)	-	-	-	-	66368
Dropout	(246, 64)	-	-	-	-	0
Add	(246, 64)	-	-	-	-	0
Layer normalization	(246, 64)	-	-	-	-	128
Time distributed	(246, 64)	-	-	-	-	4160
Time distributed	(246, 64)	-	-	-	-	4160
Dropout	(246, 64)	-	-	-	-	0
Add	(246, 64)	-	-	-	-	0
Layer normalization	(246, 64)	-	-	-	-	128
Multi head attention	(246, 64)	-	-	-	-	66368
Dropout	(246, 64)	-	-	-	-	0
Add	(246, 64)	-	-	-	-	0
Layer normalization	(246, 64)	-	-	-	-	128
Time distributed	(246, 64)	-	-	-	Relu	4160
Time distributed	(246, 64)	-	-	-	-	4160
Dropout	(246, 64)	-	-	-	-	0
Add	(246, 64)	-	-	-	-	0
Layer normalization	(246, 64)	-	-	-	-	128
Flatten	(15744)	-	-	-	-	0
Dense	(64)	-	-	-	Relu	1007680
Dense	(2)	-	-	-	Softmax	130

Priedas nr. 3

Daugiatikslio klasifikavimo modeliai

25 lentelė. Multi2 modelis

Transformacija	Dimencija	Parametrai	Sujungta su	Aktyvacijos funk.
Input layer (Skeletonas)	(990, 1)	0	-	-
Convolutional	(988, 32)	128	Input layer (Skeletonas)	Relu
Batch normalization	(988, 32)	128	Convolutional	-
Max pooling	(494, 32)	0	Batch normalization	-
Convolutional_1	(492, 64)	6208	Max pooling	Relu
Batch normalization_1	(492, 64)	256	Convolutional_1	-
Max pooling_1	(246, 64)	0	Batch normalization_1	-
Flatten	(15744)	0	Max pooling_1	-
Dense	(128)	2015360	Flatten	Relu
Batch normalization_2	(128)	512	Dense	-
Batch normalization_3	(128)	512	Dense	-
Dense_1	(32)	4128	Batch normalization_2	-
Dense_2	(32)	4128	Batch normalization_3	-
Batch normalization_4	(32)	128	Dense_1	-
Batch normalization_5	(32)	128	Dense_2	-
Dense_3	(12)	396	Batch normalization_4	-
Dense_4	(8)	264	Batch normalization_5	-
Batch normalization_6	(12)	48	Dense_3	-
Batch normalization_7	(8)	32	Dense_4	-
Dense (class_type)	(4)	52	Batch normalization_6	Softmax
Dense (correctness)	(2)	18	Batch normalization_7	Softmax

26 lentelė. Multi3 modelis

Transformacija	Dimencija	Parametrai	Sujungta su	Aktyvacijos funk.
Input Layer (Skeletonas)	(990, 1)			-
Convolutional	(988, 32)	128	Input Layer (Skeletonas)	-
Batch normalization	(988, 32)	128	Convolutional	Relu
Max pooling	(494, 32)	0	Batch normalization	-
Dropout	(494, 32)	0	Max pooling	-
Convolutional_1	(492, 64)	6208	Dropout	Relu
Batch normalization_1	(492, 64)	256	Convolutional_1	-
Max pooling_1	(246, 64)	0	Batch normalization_1	-
Dropout_1	(246, 64)	0	Max pooling_1	-
Flatten	(15744)	0	Dropout_1	-
Dense	(128)	2015360	Flatten	Relu
Batch normalization_2	(128)	512	Dense	-
Dropout_2	(128)	0	Batch normalization_2	-
Dropout_3	(128)	0	Batch normalization_2	-
Dense (class_type)	(4)	516	Dropout_2	Softmax
Dense (correctness)	(2)	258	Dropout_3	Softmax

27 lentelė. Multi4 modelis

Transformacija	Dimencija	Parametrai	Sujungta su	Aktyvacijos funk.
Input Layer (Skeletonas)	(990, 1)			-
Convolutional	(988, 32)	128	Input Layer (Skeletonas)	-
Batch normalization	(988, 32)	128	Convolutional	-
Max pooling	(494, 32)	0	Batch normalization	-
Convolutional_1	(492, 64)	6208	Max pooling	Relu
Batch normalization_1	(492, 64)	256	Convolutional_1	-
Max pooling_1	(246, 64)	0	Batch normalization_1	-
Global average pooling	(64)	0	Max pooling_1	-
Dense	(128)	8320	Global average pooling	Relu
Batch normalization_2	(128)	512	Dense	-
Batch normalization_3	(128)	512	Dense	-
Dense_1	(32)	4128	Batch normalization_2	-
Dense_2	(32)	4128	Batch normalization_3	-
Batch normalization_4	(32)	128	Dense_1	-
Batch normalization_5	(32)	128	Dense_2	-
Dense_3	(12)	396	Batch normalization_4	-
Dense_4	(8)	264	Batch normalization_5	-
Batch normalization_6	(12)	48	Dense_3	-
Batch normalization_7	(8)	32	Dense_4	-
Dense (class_type)	(4)	52	Batch normalization_6	Softmax
Dense (correctness)	(2)	18	Batch normalization_7	Softmax

28 lentelė. Multi5 modelis

Transformacija	Dimencija	Parametrai	Sujungta su	Aktyvacijos funk.
Input Layer (Skeletonas)	(990, 1)			-
Convolutional	(988, 32)	128	Input Layer (Skeletonas)	ELU
Batch normalization	(988, 32)	128	Convolutional	-
Max pooling	(494, 32)	0	Batch normalization	-
Dropout	(494, 32)	0	Max pooling	-
Convolutional_1	(492, 64)	6208	Dropout	ELU
Batch normalization_1	(492, 64)	256	Convolutional_1	-
Max pooling_1	(246, 64)	0	Batch normalization_1	-
Dropout_1	(246, 64)	0	Max pooling_1	-
Flatten	(15744)	0	Dropout_1	-
Dense	(128)	2015360	Flatten	ELU
Batch normalization_2	(128)	512	Dense	-
Dropout_2	(128)	0	Batch normalization_2	-
Dropout_3	(128)	0	Batch normalization_2	-
Dense (class_type)	(4)	516	Dropout_2	Softmax
Dense (correctness)	(2)	258	Dropout_3	Softmax