VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Detalus vaizdų panašumas naudojant trejetų tinklus

Fine-grained Images Similarity using triplet-network

Kursinis darbas

Atliko: 4 kurso 6 grupės studentas
Andrius Butkevičius (parašas)

Darbo vadovas: dr. Vytautas Valaitis (parašas)

Vilnius – 2020

TURINYS

Įvadas

Vaizdų panašumo įvertinimas ir lyginimas tampa vis plačiau naudojamas ir susilaukia vis didesnio dėmesio informacinių technologijų srityje.Visgi plėtojant šią technologiją ir siekiant išgauti kuo korektiškesnius rezultatus yra susiduriama su problemomis, nes kompiuteriui nėra lengva atskirti vizualius skirtumus ir objektų bruožų panašumus taip lengvai, kaip palyginus žmogui, kuris gali akimirskniu sugebėti atpažinti aplink jį esamus objektus. Vaizdų identifikavimas ar jų palyginimas plačiai taikomas šiuo metu, pvz.: vaizdinės tapatybės nustatymui, veido atpažinimui, esamos vietovės aptikimui pasitelkient beipiločio orlaivio užfiksuotas reljefo nuotraukas, vaizdų radimui pasitelkiant paieškos sistemas. Norint sugebėti tai atpažinti ir klasifikuoti, yra pasitelkiami įvairūs tinklų modeliai ir jie treniruojami. Tam panaudojama tokie modeliai kaip Siamo arba trejetų tinklai. Šie tinklai susideda iš dviejų ar trijų vienodų ir lygiagrečių konvoliucinių neurono tinkle esančių atšakų, kurios tarpusavyje dalinasi svoriais, kurių dėka galima gauti aukšto lygio nuotraukų bruožų atvaizdavimą taip leidžiant panašioms nuotraukoms būti kuo arčiau sujungtoms viena su kita funkcijų erdvėje. Tuo metu nepanašiems vaizdams – būnant kuo toliau nuo teisingų nuotraukų.
Rinkoje yra siūlomi įvairūs konvoliucinių neuronų tinklų modeliai, pvz.: Alexnet, VGGNet, GoogLetNet, ResNet [VMJ19]. Verta paminėti, kad visi šie modeliai turėjo efektyvius atpažinimo rezultatus ILSVRC. Taigi vis dažniau yra pasitelkiama prieš tai minėti trejetų tinklai, kurie optimizuoja bruožų atstumus erdvėje. Visgi sudėtingiausia problema išlieka dėl semantinio tarpo problemos, kuri atsiranda tarp žemos rezoliucijos nuotraukos pikselių užfiksuotų kompiuterinių sistemų ir aukšo lygio semantinių koncpetų, kurias suvokia žmonės. Dėl to reikia rasti geresnių būdų, kaip sugebėti pateikti nuotrauką kompiuteriui ir gauti gilesnias jos semantinius bruožus.
Todėl šio darbo tikslas yra pasiūlyti pasrinktą trejetų tinklų modelį [VMJ19], kuris sugebėtų atpažinti žmonių siluetus su pasirinktu duomenų rinkiniu. Palyginti pasirinktą modelį su kitais esamais rinkoje modeliais, palyginti rezultatus. Taip pat išskirti metrikas tam. Surasti su kuriais vaizdais pasirinktas modelis prastai vykdo atpažinimą ir kodėl.

Uždaviniai

1. Ištreniruoti trejetų tinklų modelį atpažįstant žmonių siluetus, nufotografuotus viešoje vietoje.
2. Išskirti metrikas pagal kurias galėtų analizuoti gautus rezultatus, panaudojant trejetų tinklų modelį.
3. Išskirti pasirinkto tyrimo rezultatų trūkumus ir išsiaiškinti galimas sritis ateities darbams.
4. Palyginti kitų neuronų tinklų modelių architektūras.

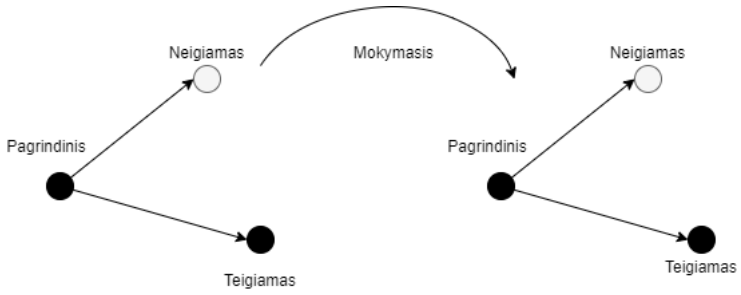
1. Trejetų ir Siamo tinklų tyrimų literatūros analizė

1.1. Gilieji konvoliuciniai neuronų tinklai

Giliajame treniravime(mokyme), šis tinklas yra klasė iš giliųjų neuronų tinklų, kuri dažniausiai naudojama pritaikant vaizdų atpažinimą. Tai yra algoritmas, kuris įvestyje pasiima vaidą (nuotrauką ar paveiksliuką), priskiria jam svorius ir jos bruožų tendencijas įvairiose įvesties dalyse ir sugeba pagal tai atskirti panašumus lyginant su kitomis įvestimis.

1.2. Trejetų tinklai

Trejetų tinklų modelis būvo pasiulytas 2014m., moksliniko Chang Wang [WSL+14]. Modelio veikimo principas susideda iš trijų identiškų konvoliucinių neuroninių tinklų šakų, kurie tapusavyje dalijasi gautais svoriais. Kai trejetų tinklas įvestyje gauna tris pavyzdžius, išvestyje yra grąžinama dvejos tarpinės reikšmės, kurios nurodo atstumus lyginamus su pagrdine įvestimi ir kitomis dvejomis įvestimis. Jų rezultatas yra vektoriai. 1pav. parodytas abstraktus treniravimosi procesas, trejetų tinklų modelio.



1 pav. Trejeto tinklo treniravimosi procesas [YLL+19]

1.2.1. Architektūra

Trejetų tinklų modelis įvedimo dalyje reikalauja trijų įvesčių tuo pačiu metu. Kiekvienas jų turi savo pavadinimą ir reikšmę tinkle, pagrindinis x^a, pozityvus x^p, neigiamas x^n. Įvesties poros x^a ir x^p yra tos pačios kategorijos arba panašūs semantiškai įvestys. Tuo metu x^a ir x^n yra skirtingos kategorijos arba nepanašūs vaizdai. Pasitelkiant funkcijas yra apskaičiuojama jų semantiniai panašumai atstumo erdvėje (nes jie yra pateikiami kaip vektoriai). Ši funkcija yra vadinama nuostolių funkcija, jos galima funkcija aprašyta žemiau. Kur parametras α rodo tarpą tarp x^a ir x^p bei x^a ir x^n. N reiškia skaičių nusakantį kiek trejetų įvesčių yra pateikiama. Tikslas yra pasiekti, kad x^a ir x^p būtų mažiau nei x^a ir x^n [SKP15].

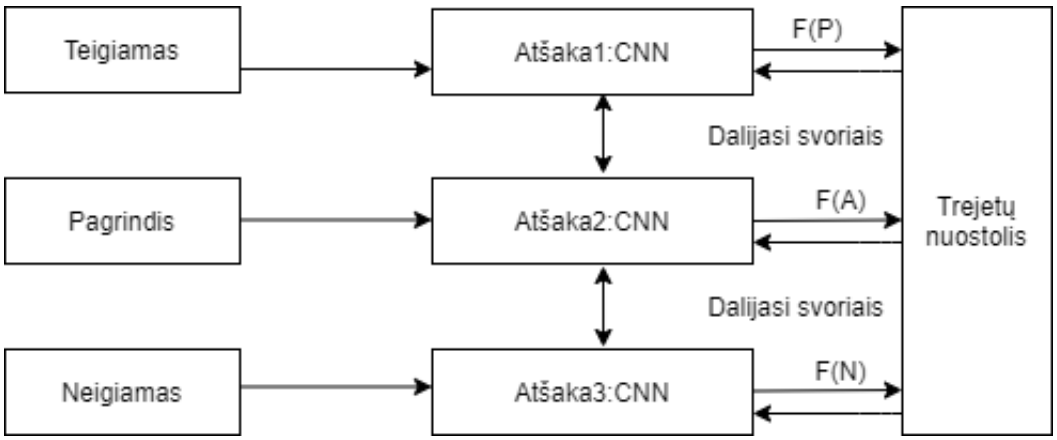
$$loss = \sum_{x=1}^{N} max(d(x^a, x^p) - d(x^a, x^n) + \alpha, 0)$$

Kur d yra atstumo metrika, taip kad $d(x^a, x^p) < d(x^a, x^n)$, o N yra skaičius trejetų.

Nuostolių funkcija nusako, kaip gerai algoritmas modeliuoja pasirinktą duomenų rinkinį. Jeigu prognozė rezultatui yra netiksli, ši funkcija grąžina didesnę reikšmę. Neuroninių tinklų modeliams galima naudoti ir kitas nuostolių funkcijas, priklausomai nuo duomenų rinkinio ir norimo rezultato gauti.

Alternatyvios nuostolių funkcijos. Artima šiai funkcijai yra ši lygtis. Ji naudoja kvadratinį Euklido [VMJ19] atstumą kaip atstumo metriką ir $\alpha > 0$, nurodanti ribą tarp dviejų vaizdų ir absoliučios sumos reikšmės.

$$loss = \sum_{x=1}^{N} [||x^a - x^p|| - ||x^a - x^n|| + \alpha]$$



2 pav. Trejeto tinklų veikimo principas kartu su neuroninių tinklų lygiagrečiomis atšakomis

Šis neuroninių tinklų architektūrinis sprendimas padeda apmokyti duomenų klasifikavimą išskaidant duomenis pagal panašumus ir skirtumus (2 pav.). Jų metu keli lygiagretūs gileiji neuroniniai tinklai yra apmokami bei tuo pačiu metu jie dalijasi svoriais vieni su kitu treniruojant tinklą. Trejetų tinkle tikslas yra sukurti trejetus, kurie susideda iš pagrindinio p^a, teigiamio x^p, neigiamio

p^n – įvesties elementų. Pagrindinis elementas, tai kažkokia įvestis (nuotrauka, paveiksliukas, muzikos įrašas ir t.t), kuriai mes bandome rasti atitikimą iš kitos įvesties. Teigiama įvestis – panašus elementas atitinkamai pagrindiniai nuotraukai. Tuo tarpu neigiamas elementas skaitosi tas, kuris neturi panašumų su pagrindine nuotrauka. Neuroninai tinklai apskaičiuoja $\pi : f(\pi) \in$ Šie trys elementai yra įvedami nepriklausomai vienas nuo kito į tris identiškus giliuosius neuroninius tinklus, kurie dalijasi vienoda architektūra ir parametrais. Jų metu yra apskaičiuojami atstumai tarp šių elementų. Šiam apskaičiavimui yra naudojama kaip ir prieš tai minėta nuostolių funkcija.
Įmanomi rezultatai trjetų tinkle:

- Lengvas trejatas: trejetai, kurie turi nuostolį, su reikšme 0, nes $d(x^a, x^p) + \alpha < d(x^a, x^p)$
- Sudėtingas trejetas: trejetai, kurių neigiamumas yra arčiau pagrindinio negu teigiamo, $d(x^a, x^p) < d(x^a, x^p)$
- Pusiau sudėtingas trejetas: trejetai, kurių neigiamumas nėra arčiau teigiamo, tačiau vis tiek turi teigiamą nuostolį: $d(x^a, x^n) < d(x^a, x^p) + \alpha$

1.3. Siamo tinklas

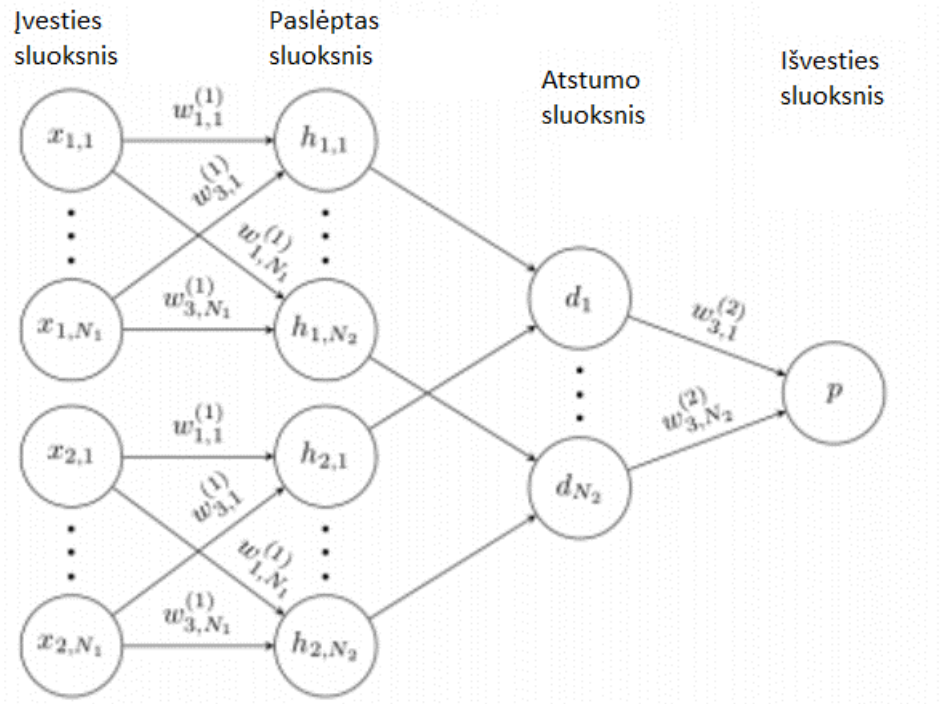
Pirma kartą publikuotas 1990 metais, autorių Bromely ir Yann LeCun, siekiant išspręsti parašo verifikavimo problema kaip vaizdo atitikimo probelmą [BGL+90].

1.3.1. Architektūra

Tinklo modelis – neuronų tinklas turintys kelis ar daugiau identiškus neuroninių tinkle atšakas su vienodais parameterais [MKR16] (panašiai kaip ir trejetų tinklų modelyje). Šis modelis naudoja vienodus svorius vykdant tuo pačiu metu ir panaudojant du skirtingus įvesties vektorius, apskaičiuojant palyginamus išvesties vektorius. Šie vienodi tinklai su skiringais įvesties duomenimis yra sujungti pagal energijos funkciją E. Ši matematinė funkcija apdoroja kai kurias metrikas tarp labiausiai išryškintų bruožų vaizdavimų kiekvienoje pusėje. Parametrai tarp šių vienodų tinkle yra apriboti, kad būtų vienodi. Svorių apjungimas užtikrina, kad du labai panašūs vaizdai nebūtų tarp jų atitinkamų tinkle atšakų išvesti labai skirtinguose erdvės lokacijose. Taip pat, reikia paminėti, kad tinklas yra simetriškas, nesvarbu kuriam iš vienodų tinklų įvesime vaizdą, visada gausime tokias pačias metrikas. Standartinė išlaidų funkcija skirta mokymosi pavyzdžiui (x_1, x_2) yra pasiūlyta Hadselio.

$$L(W,(Y, x_1, x_2)) = 0.5(1–Y)(D_w)^2 + (Y)0.5max(0, m–D_w)^2$$

, jei (x_1, x_2) yra panašios poros, ir $Y = 1$ kitu atveju. m yra riba nurodanti norimą slenkstį atstumui tarp x_1 ir x_2 jeigu jie nėra panašūs. Dėl laisvo suregualiavimo, dažniausiai taip yra sunkiau ištreniruoti trejetų tinklus negu Siamo tinkle. Tarkime x_1, x_2 yra pora vektorių, Y laikysime dvejatainę žymę, kur $Y = 1$ reiškia, kad vektoriai x_1, x_2 yra laikomi panašiais, $Y = 0$ - priešingu atveju. D_w - parametrizuota atstumo funkcija (Euklido).



3 pav. Siamo tiklų modelis

1.4. Nefiksuoto dydžio vaizdai konvoliuciniuose tinkluose

Gilieji konvoliuciniai tinklai reikalauja fiksuoto dydžio įvesties vaizdų. Tačiau realiame gyvenime nuotraukų ar paveiksliukų dydžiai nėra fiksuoti ir varijuoja plačiai.

Jeigu yra bandoma primiktynai pakeisti į reikiama dydį, nuotrauką karpant ar deformuojant, informacija, patalpinta nuotraukose bus prarasta. To pasekoje tikslumas nuotraukų klasifikacijos ar objektų identifikavime bus sumažintas ir nepataisomai sugadintas. Nors neuroniniuose tinkluose, konvoliuciniai sluoksniai nereikaljau fiksuoto dydžio įvesties ir geba generuoti specifinius bruožus vaizdo iš bet kokio dydžio nuotraukų. Visgi, pilnai sujungti sluoksniai privalo turėti fiksuoto dydžio įvestį dėl jų pačių apibrėžimo. Dėl to apribojimas fiksuoto dydžio nuotraukų ateina tik iš pilnai sujungtų sluoksnių reikiamos ypatybės. Viena iš šios problemos sprendimo būdų yra naudoti erdvinės piramidės talpinimą [HZR+15], tokiu būdų galima bandyti identifikuoti vaizdus, kurių rezoliucijos yra skirtingos.

Šis būdas ištraukia vaizdo bruožus iš bruožų žemėlapio (angl. map) per $4x4$, $2x2$ ir $1x1$ kvadratų tinklelio. Tada SPP sluoksnis pateikia $16+4+1 = 21$ skirtingus aruodus (angl. bin) ir gauna fiksuoto dydžio išvestį kviečiant kviekvieną bloką. Po erdvinės piramidės talpinimo būdo išgavimo, bet kuris bruožų žemėlapis gali generuoti 5736 dimensijų ypatybių vektorius, kur $5736 = 24x26$. SPP sluoksnis pasiima ypatybes ir generuoja fiksuoto dydžio išvestis, kuris galiausiai yra perduodamas į pilnai sujungtus sluoksnius.

2. Įvertinimo funkcija

Kelios įvertinimų metrikos yra naudojamos: panašumo tikslumas bei *score-at-top-K*, kai $K = 30$. Panašumo tikslumas yra išreiškiamas procentaliai pagal tai, kiek trejetų buvo korektiškai sureitinguota. Sakykime, kad turime trejetą, su šiais įvesties parametrais $tt = (x^a, x^p, x^n)$, kur x^p turėtų būti arčiau(panašesnis)šalia x^a. Laikant, kad x^a yra įvesties užklaus,a žiūrime į rezultatus kitų įvesties duomenų. Jeigu x^p yra reitinguojamas aukščiau nei x^n, tai tada teigiame, kad atitinkamas trejetas yra sureitinguotas teisingai. Kita mums reikalinga metrika, kuri buvo užsiminta anksčiau *score-at-top-K*. Jis nusako skaičių teisingai sureitinguotų trejetų bei atimant iš jo skaičių, kuris nusako neteisingai sureitinguotus trejetus iš pogrupio trejetų, kurių reitingas yra didesnis, nei kintamasis K. Pogrupis yra pasirenkamas tokia tvarka: kiekvienam užklausos paveikslėliui iš duomenų aibės, ištraukia 1000 naujų paveikslėlių iš tos pačios teksto užklausos ir yra bandoma taip reitinguoti juos, pasitelkiant išmoktas metrikas. Jei trejetų reitingas yra aukštesnis negu K, jeigu jo x^p arba x^n tada yra tarp geriausiai reitinguojamų K kaimynų iš užklausos su paveikslėliais x^a.

2.1. Pasirenkami duomenis įvertinimui analizuoti

Kadangi darbo tema susijusi su detalių vaizdų panašumu, kuris negali būti charakterizuotas pagal paveikslėlių žymes, buvo panaudota trejetų duomenų rinkinys įvertinant vaizdo panašumus modeliams. Buvo paimta 1000 populiariausių teksto užklausų su atrinktais trejetais (x^a, x^p, x^n) su Google 50 paieškos rezultatų kiekvienai užklausai [WSL+14].

Efektyvumas yra nurodomas pirmoje lentelėje. ”DeepRanking” parodytas lentelėje yra gilusis reitingavimo modelis treniruotas su 20

Method	Precision	Score-30
ConvNet	82.8%	5772
Single-scale Ranking	84.6%	6245
OASIS on Single-scale Ranking	82.5%	6263
Single-Scale & Visual Feature	84.1%	6765
DeepRanking	**85.7%**	**7004**

4 pav. anašumo tikslumo metrikų rezultatai [WSL+14]

Method	Precision	Score-30
Wavelet [9]	62.2%	2735
Color	62.3%	2935
SIFT-like [17]	65.5%	2863
Fisher [20]	67.2%	3064
HOG [4]	68.4%	3099
SPMKtexton1024max [16]	66.5%	3556
L1HashKPCA [14]	76.2%	6356
OASIS [3]	79.2%	6813
Golden Features	80.3%	**7165**
DeepRanking	**85.7%**	7004

5 pav. Panašumo tikslumo metrikų rezultatai [WSL+14]

3. Trejetų tinklo vaizdų atpažinimo tyrimas ir vertinimas

3.1. Kaip išspręsti semantines vazidų skirtumo problemas

Norint pagerinti atsirandančias vaizdų problemas, aprašytas viršuje

3.2. Treniravimo ir testavimo aplinka

Trejetų tinklo modelis buvo treniruoamas naudojant Lenovo Y-700 kompiuterį su šiomis specifikacijomis: CPU i7 6700K (4 branduolių), 8GB RAM, GPU Nvidia GTX 960m (4GB RAM). Norint pasileisti trejetų modelį būtent šiam kompiuteriui (priklausomai nuo mašinos tai gali skirtis) buvo atsisiųsta Nvidia CUDA programinė įranga su tikslu, kad modelio treniravimas būtų vykdomas pasitelkiant vaizdo plokštę, o ne procesorių.
Priežastis kodėl buvo pasirinkta vaizdo plokštė yra todėl, kad gilusis mokymasis yra intensyvi skaičiavimo užduotis. Į gilųjį mokymasi įeina didžiuliai matricų skaičiavimai (ypač sandauga) ir kitos operacijos, kurios gali veikti paraleliai todėl vaizdo plokštė ateina į pagalbą, nes viena vaizdo plokštė gali turėti tūkstančius branduolių, tuo metu procesorius turi žymiai mažiau branduolių, nors jie ir žymiai greitesni negu vaizdo plokštės [KK18].
Pasirenktas trejetų tinklų modelis yra implementuotas naudojant Tensorflow karkasą, Python 3.5.1. Vaizdų duomenų rinkinyje buvo 3884 nuotraukos, kuriose jau buvo surikiuotos teisinga eilės tvarka, kur pirma nuotrauka pagrindinė, antra nuotrauka - teigiama (kuri yra panaši į pagrindinę) bei neigiama (nepanaši į pagrindinę nuotrauką) ir tokia eilės tvarka yra išsidėstę visos kitos nuotraukos. Ištreniravus šiuos duomenis yra bandoma analizuoti modelio tikslumą imant vaizdų pavyzdžius iš testinio duomenų rinkinio.

3.3. VGG16 tinklo modelis

Panaudota architektūra naudoja VGG16 tinklo bazės sluoksnius klasifikavimui todėl tik labai panašūs vaizdai gali būti naudojami duomenų treniravimui tam, kad pagerinti tinklo atlikimą bei išsaugoti skaičiavimo išteklius.
VGG16 yra konvoliucinių neuroninių tinklų modelis, išrastas K. Simonian ir A. Zisserman (Oksfordo universitetas). Modelis sugeba pasiekti 92.7% top-5 testų tikslumą (treniravimo duomenys pasiimti iš ImageNet). VGG16 buvo treniruojamas savaitę iš savaitės naudojant Nvidia Titan Black vaizdo plokščių šeimas.

3.4. Tyrimo rezultatai

Šiame darbe buvo bandyta atpažinti detalius vaizdus, naudojant vaizdų duomenų rinkinį CUHK01. Šiame duomenų rinkinyje yra pavaizduota 3884 viešose vietose einančių pėsčiųjų kadrai. Todėl naudojant ši duomenų rinkinį buvo analizuojama, kaip su pasirinktu trejetų modeliu seksis atpažinti skirtingus žmogaus kadrus, kurie buvo nufotografuoti įvairiai - iš kelių pusių, iš

nugaros ir priekio, esant skiringai žmogaus eisenos pozicijai, skirtingam apšvietimui ir kontrastamas, įsiterpiant kitiems objektams prie pėsčiojo.
Ištreniravus trejetų tinklą su pasirinktu duomenų rinkiniu ir pradėjus testuoti jo rezultatus, gavau 85,7% tikslumo rezultatą. Tai yra gan aukštas įvertinimas, nes atpažinti to paties pėsčiojo skirtingus kadras nėra lengviausia užduotis, nes susiduriame su kliūtimis, aprašytomis anksčiau. Jeigu palyginsime kaip sekėsi kitiems neuroninių tinklų modeliams atpažinti būtent šį duomenų rinkinį, turime tokius rezultatus: 90.4% , kur neuronų tinklų modelį sukūrė mokslininkai iš Kinijos Mokslų ir Technologijų universiteto. [ZLZ+19]

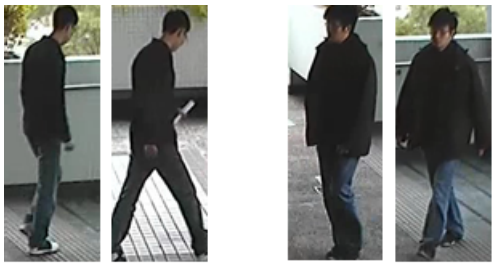
6 pav. Žmogaus skiringi eisenos kadrai

Taip pat, žemiau pateikiu kitus rezultatus, gautus su kitais neuroniais tinklų modeliais:

Spindle	94.4%
PSE	86.6%
Part-Aligned	94.4%
AACN	96.7%

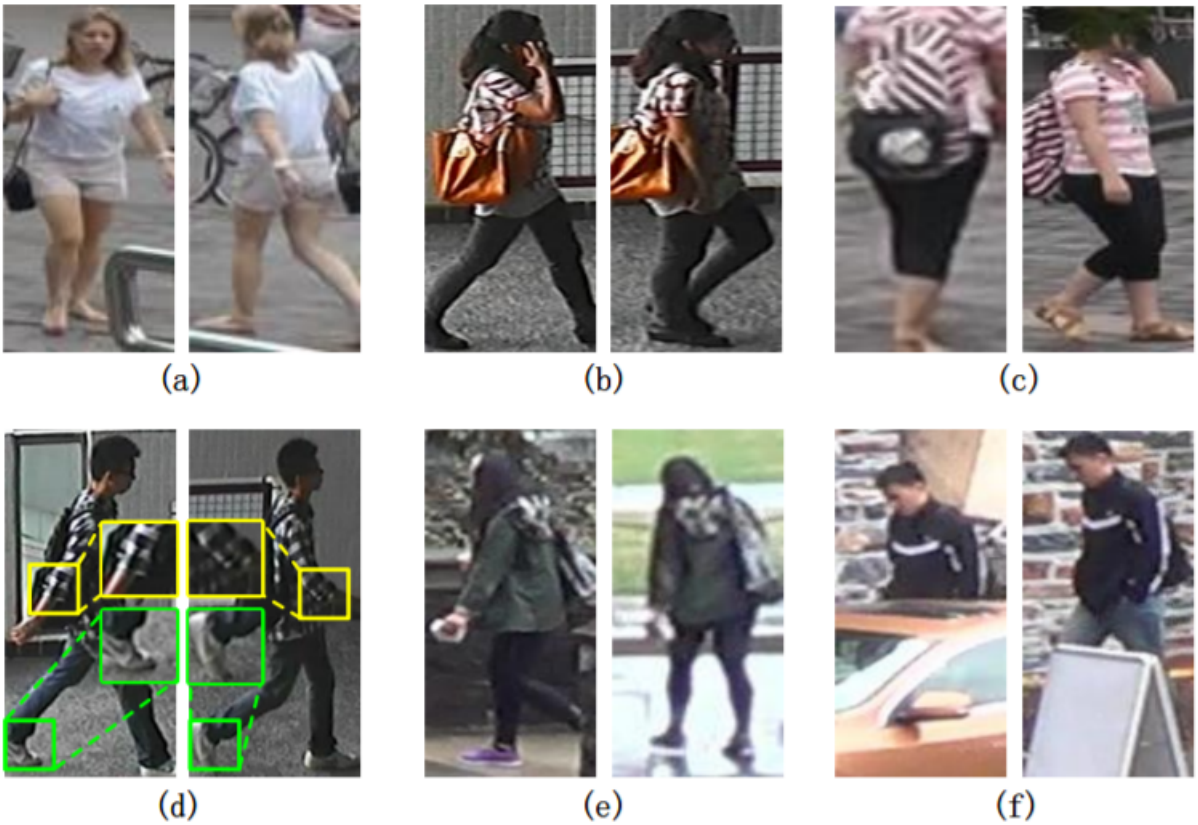
3.5. Vaizdai su prastais rezultatų palyginimais

Visgi ne visi vaizdai buvo tesingai palyginimi ir nustatomi trejetų tinklo modelio. Todėl šiame darbe buvo ieškoma ir nagrinėjama, kurios vaizdų įvestis turėdavo prastus rezultatus bei buvo bandoma išsiaiškinti kas sukeldavo prastus rezultatus.



7 pav. Blogai identifikuotos vaizdų poros

Dauguma nuotraukų, kurios buvo blogai identifikuotos skirėsi viena nuo kitos spalvų intensivumu, skirtingomis spalvomis. Pavyzdžiui matome, kad tiek pirmoje ir antroje nuotraukų porose įsiterpusi nauja žalia spalva (žalumos objektai) leidžia manyti, kad tai yra viena iš priežasčių, kodėl nepavyko tinkamai identifikuoti to paties asmens kadrus. Taip pat antroje poroje matome, kad žmogaus judėjimo kampas į objektyvą skiriasi, o ir nuotraukų šviesos intensyvumas yra kitoks. Todėl tai leidžia daryti prielaida, kad ir vaizdų palyginimo rezultatai bus prasti, jei nuotraukos bus nufotografuotos ne kokybiškai - apšvietimo skirtumai, netvarkingas fonas ar atsiradusi okliuzija [ZLZ+19].



8 pav. Sunkiai identifikuojamų nuotraukų poros

Viršuje esančios žmonių nuotraukų poros yra sudėtingai identifikuojamos dėl įvairių priežasčių: a) dėl skirtngo objektyvo kampo, b) skirtingos žmogaus pozos, c) prastos nuotraukos kadro užfiksavimo (susiliejas vaizdas), d) ne vienodame lygyje esančios žmogaus kūno dalis, atsiradusios dėl jo judėjimo, e) netvarkingas fonas, f) okliuzija.

3.6. Sprendimo būdai

Visgi ieškant literatūroje informacijos, kurioje būtų aprašyti sprendimo būdai, kai lyginami vaizdai vienas su kitu yra nekorektiški bruožų atžvilgiu, tačiau panašūs ar priklauso tai pačiai klasei, yra siūlomi keli sprendimų būdai. Pavyzdžiui, ypatybių fokusavimą nukreipiant į lokalias detales buvo sukurtas nesudėtingas suskirstymas asmens nuotraukos į kelias fiksuotas ir nekintančias dalis, pavyzdžiui į horizontales juosteles ir taip mokantis lokalių nuotraukų bruožų atpažinimo. Visgi, kaip tyrimai iš literatūros nurodė toks skirstymas prastai lygina identifikuojamo asmens kūno dalis [ZLZ+19]. Kituose tyrimuose ir bandymuose buvo bandoma naudoti žmogaus pozą, tokiu būdu identifikuojant skirtingas kūno dalis: kojas, rankas, veidą. Tačiau ir toks atpažinimo metodas yra prastas, kad gauti norimus rezultatus, nes identifiuoti tas pačias žmogaus kūno dalis iš skirtingų žmogaus padėčių kelia problemų dėl asmens skirtingos judėjimo pozos nuotraukos [ZLZ+19].

4. Rezultatai ir išvados

4.1. Rezultatai

1. Trejetų tinklai palyginti su Siamo modeliu, išskirti trūkumai ir privalumai.
2. Išskirti esamų tyrimų trūkumai
3. Aprašyti trjetų tinklų pasirinktas architektūros modelis.

4.2. Išvados

1. Yra sudėtinga lyginti trejetų tinklų ir Siamo modelius, nes jų architektūra gali varijuota pagal tai kaip modelis yra įgyvendintas. Teisingiau yra lyginti specialius jų sukurtus modelius vienas su kitu.
2. Pasirinktas trejetų tinklų modelis parodė neblogus rezultatus identifikuojant žmonių eisenos kadrus.
3. Identifikuojant žmones nuotraukose susiduriama su rimtomis problemomis, kai to paties žmogaus nuotraukos skiriasi dėl pašalinių priežasčių. Kas lemia kad žmonių identifikacija yra sudėtingas procesas, kuris tam tikroms nuotraukoms neturi sprendimo būdų.

4.3. Ateities darbai

1. Išbandyti kitą trejetų modelį su pasiūlytais būdais, kurie gebėtų dar geriau atpažinti žmonių siluetus
2. Atlikti atpažinimo rezultatų analizę su kitis neuronų tinklais.
3. Ištestuoti su daugiau duomenų rinkinių. Stebėti ir lyginti rezultatus
4. Iškelti savo palyginimo metodą, kuris gebėtų teisingai identifikuoti žmones nuotraukose.

5. Priedai

5.1. Žodynas

- Svoris(angl. weight) - parametras neuroniname tinkle, kuris transformuoja įvesties duomenys paslėptuose sluoksniuose.
- Nuostolių funkcija(angl. loss function) - funkcija, padedanti optimizuoti svorius, taip sumažinant neatitikimų nuostolius.
- Detalieji vaizdai(angl. fine-grained) - detalūs vaizdai. Vaizdo klasifikavimo užduotyse, tai yra įvesties vaizdai, kuriuos yra sudėtinga išskirti klasėms, pvz.: identifikuojant skirtingų markių automobilius.
- Aktyvacijos funkcija(angl. activation function) - funkcija, skirta nustatyti ar neuronas turi būti aktyvuotas, skaičiuojant svorių sumą bei pridedant postūmio parametrą.
- Postūmis(angl. bias) - papildomas parametras neuroniniuose tinkluose, kuris padeda koreguoti išvesti kartu su svorių įvesties suma, skirta neuronams perduoti. Taip pat šis parametras leidžia perstūmti aktyvacijos sumą nuo kairės į dešinę.

Literatūra

[BGL+90] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger ir Roopak Shah. Signature verification using a siamese time delay neural network, 1990.

[HZR+15] Kaiming He, Xiangyu Zhang, Shaoqing Ren ir Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition, 2015.

[YLL+19] Xinpan Yuan, Qunfeng Liu, Jun Long, Lei Hu ir Yulou Wang. Deep image similarity measurement based on the improved triplet network with spatial pyramid pooling, 2019.

[KK18] Amr Kayid ir Yasmeen Khaled. Performance of cpus/gpus for deep learning workloads, 2018.

[MKR16] Iaroslav Melekhov, Juho Kannala ir Esa Rahtu. Siamese network features for image matching, 2016.

[SKP15] Florian Schroff, Dmitry Kalenichenko ir James Philbin. Facenet: a unified embedding for face recognition and clustering, 2015.

[VMJ19] Vytautas Valaitis, Virginijus Marcinkevicius ir Rokas Jurevicius. Learning aerial image similarity using triplet networks, 2019.

[WSL+14] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen ir Ying Wu1. Learning fine-grained image similarity with deep ranking, 2014.

[ZLZ+19] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng ir Zhibo Chen. Densely semantically aligned person re-identification, 2019.