

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS KATEDRA

3D objektų atpažinimas iš 2D nuotraukų

3D object recognition from 2D images

Magistro baigiamasis darbas

Atliko:	Aleksas Vaitulevičius	(parašas)
Darbo vadovas:	prof. habil. dr. Olga Kurasova	(parašas)
Recenzentas:	doc. dr. Vardauskas Pavardauskas	(parašas)

Vilnius – 2020

Santrauka

Glaustai aprašomas darbo turinys, pristatoma nagrinėta problema ir padarytos išvados. Santraukos apimtis ne didesnė nei 0,5 puslapio. Santraukų gale nurodomi darbo raktiniai žodžiai.

Raktiniai žodžiai: Klasifikavimo uždavinys, 3D objektai, dirbtiniai neuroniniai tinklai, kapsuliniai neuroniniai tinklai, Konvoliuciniai neuroniniai tinklai

Summary

Santrauka anglų kalba.

Keywords: classification, 3D objects, artificial neural networks, CapsNet neural networks, convolutional neural networks

Turinys

Ivadas	4
1. 3D objektų atpažinimo iš 2D nuotraukų uždavinys, jo sprendinių pavyzdžiai	7
2. Konvoliucinių dirbtinių neuroninių tinklų apžvalga	9
3. Kapsulinių dirbtinių neuroninių tinklų apžvalga ir kalibracija	10
4. Kapsulinių ir konvoliucinių neuroninių tinklų eksperimentiniai tyrimai	11
Rezultatai ir išvados	12
Literatūra	13
Santrumpos	14
Priedas Nr.1	
Priedas Nr.2	

Ivadas

Vienas iš fundamentalių kompiuterinės regos uždavinių yra informacijos apie 3 dimensijų (3D) pasaulį išgavimas naudojant 2 dimensijų (2D) nuotraukas. Šio uždavinio tikslas yra atpažinti konkrečius 3D objektus, naudojant jų, skirtingų apžvalgos taškų 2D nuotraukas. Šiam tikslui pasiekti, yra konstruojami objektų atpažinimo algoritmai, kurie klasifikuoja 2D nuotraukas į klases, kurios atstovauja vieną iš 3D objektų modelių.

3D objektų atpažinimas iš 2D nuotraukų yra naudojamas srityse, kuriose turimi 3D objektai turi būti atpažinti iš visų galimų 2D nuotraukų, turint tik poaibį šių nuotraukų. Keli šių sričių pavyzdžiai yra vogtų objektų aptikimas - turint algoritmą, atpažįstantį konkretų automobilį, galima iš viešų erdvių nuotraukų atrasti automobilio poziciją, vietos nustatymas - turint algoritmą, atpažįstantį konkrečius objektus esančius skirtingose vietovėse, ir tų vietovių koordinatės galima nustatyti kurioje vietovėje buvo padaryta nuotrauka. Deja, laiko ir duomenų kaštai yra per dideli, kad pasiekti absoliutų tikslumą sprendžiant šį uždavinį. Todėl taikomi metodai yra euristiniai. Dėl to renkantis metodą, spręsti 3D objektų atpažinimo iš 2D nuotraukų uždaviniui, reikia atsižvelgti į laiko kaštus ir kaip tiksliai tuo metodu pagrįstas algoritmas klasifikuoja 2D nuotraukas, sprendžiamas šį uždavinį. Šiame darbe, bus atliekami tyrimai, skirti nustatyti metodą, kuris būtų pranašesnis sprendžiant 3D objektų atpažinimo iš 2D nuotraukų uždavinį. Pasirinkti metodai yra konvoliucinių ir kapsulinių dirbtinių neuroninių tinklų architektūros.

Gana dažnai naudojamas metodas, šiam uždaviniui spręsti, yra dirbtiniai gilieji neuroniniai tinklai. 3D objektų atpažinimo iš 2D nuotraukų uždavinyje, naudojami duomenys - 2D nuotraukos, yra nestruktūrizuoti, jiems sudėtinga vykdyti požymių išgavimą. Todėl daugelis kitų sprendimų nėra tokie patrauklūs kaip dirbtiniai gilieji neuroniniai tinklai, dėl savo sugebėjimo efektyviai vykdyti automatinį požymių išgavimą iš nestruktūrizuotų duomenų. Tačiau, norint pasiekti aukštą klasifikavimo tikslumą, naudojant šį metodą, yra reikalingas didelis kiekis duomenų. Konkrečiai šiam uždaviniui reikia didelio kiekio 3D modelių. Laimei, šiuo metu egzistuoja viešai prieinamos didelės 3D repozitorijos. Tokios kaip 3D Warehouse, TurboSquid, ir Shapeways. Dėl to, šiuo metu daugelis architektūrų jau yra išbandytos sprendžiant 3D objektų atpažinimo iš 2D nuotraukų uždavinį. Pavyzdžiui viena iš architektūrų, kuri buvo išbandyta, yra konvoliucinio gilaus pasitikėjimo neuroninio tinklo architektūra. Šiai architektūrai atlikti tyrimai yra aprašyti darbe [WSK⁺15].

Šiuo metu, šiam uždaviniui spręsti, optimaliausius rezultatus, laiko ir tikslumo atžvilgiu, pasiekusi dirbtinio neuroninio tinklo architektūra yra konvoliuciniai neuroniniai tinklai. Tyrimai, kuriuose ši architektūra buvo išbandyta, yra aprašyta darbuose [Mas18; SMK⁺15]. Darbe [WSK⁺15] atlikto tyrimo rezultatai parodo, kad pateiktas sprendimas, kuriame 3D objektų atpažinimas yra konstruojamas naudojantis tik 2D nuotraukomis, yra tikslesnis 8% (77% → 85%). Todėl šiame darbe bus daromi eksperimentai su dirbtinio neuroninio tinklo architektūrų įgyvendinimais, kurie yra pagrįsti būtent šiuo metodu. Tad šiam darbui, vienas iš pasirinktų metodų yra konvoliucinis neuroninis tinklas, kurio įgyvendinimas ir tyrimai yra aprašyti darbe [SMK⁺15]. Mat šio darbo įgyvendinimas naudoja tik 2D nuotraukas, konstruojant 3D objektų atpažinimo algoritmą.

Kita tiriama dirbtinio neuroninio tinklo architektūra yra kapsuliniai neuroniniai tinklai. Lyginant su konvoliuciniais neuroniniais tinklais, tai gana nauja architektūra. Aprašyta 2017 metais

darbe [nNH17] Kapsulinių neuroninių tinklų architektūros veikimo principas tiksliau imituoja žmogaus regą, remiantis faktu, kad žmogaus regą ignoruoja nereikšmingas vaizdo detales, naudodama tik seką fokusuotų taškų, taip apdorojama tik dalį vaizdinės informacijos su labai aukšta rezoliucija. [nNH17] darbe atliktas tyrimas parodo, kad ši architektūra atlieka skaičių klasifikavimo užduotį tiksliau nei konvoliucinis neuroninis tinklas. Kitas tyrimas, kuris yra atliktas darbe [MC17] su 4 duomenų rinkiniais, kuriuose yra veidai, kelio ženklai ir kasdienius objektus, parodo, kad dabartiniai kapsuliniai neuroniniai tinklai nevisą laiką yra pranašesni už konvoliucinius neuroninius tinklus. Parinkus geresnius parametrus arba sukūrus geresnį dizainą, konvoliucinis tinklas dar vis būna pranašesnis už kapsulinį neuroninį tinklą. Tačiau taip pat šiame darbe yra teigiama, kad kapsuliniai neuroniniai tinklai dar nėra pasiekę pilno savo potencialo ir tolimesni tyrimai turi būti atlikti.

Tad šiame darbe yra keliama tokia **hipotezė**:

Kapsuliniai neuroniniai tinklai sprendžia 3D objektų atpažinimo iš 2D nuotraukų uždavinį efektyviau nei konvoliuciniai neuroniniai tinklai remiantis apmokymo laiko ir tikslumo kriterijais.

Tikimasi, kad, sprendžiant 3D objektų atpažinimo iš 2D nuotraukų uždavinį, kapsulinio neuroninio tinklo mokymas truks trumpiau nei konvoliucinio neuroninio tinklo. Taip pat, kad apmokytas kapsulinis neuroninis tinklas vykdys klasifikavimą tiksliau nei konvoliucinis neuroninis tinklas.

Siekiant patikrinti iškeltą hipotezę reikia atlikti šiuos uždavinius:

1. Išanalizuoti ir nustatyti dabartinį efektyviausią 3D objektų atpažinimo iš 2D nuotraukų uždavinio sprendinį, remiantis literatūra.
2. Išanalizuoti kapsulinių neuroninių tinklų veikimą, remiantis literatūra.
3. Surasti duomenis, skirtus spręsti 3D objektų atpažinimo iš 2D nuotraukų uždaviniui.
4. Eksperimentiškai nustatyti efektyviausius parametrus ir konfigūracijas, skirtus spręsti 3D objektų atpažinimo iš 2D nuotraukų uždaviniui, kapsulinio neuroninio tinklo implementacijai, remiantis apmokymo laiko ir tikslumo kriterijais.
5. Eksperimentiškai palyginti 1. užduoties rezultatą su 3. užduoties.

Šiame darbe atlikta:

1. Nustatyta, kad šiuo metu efektyviausias 3D objektų atpažinimo iš 2D nuotraukų uždavinio sprendinys yra konvoliuciniai neuroniniai tinklai, lyginant eksperimentų, aprašytų skirtinguose literatūros šaltiniuose, rezultatus. Šiuose šaltiniuose buvo surasta konvoliucinio neuroninio tinklo implementaciją ir duomenys skirti apmokymui ir testavimui.
2. Išanalizuotas kapsulinių neuroninių tinklų veikimas, surasta jo implementacija.
3. Eksperimentiškai nustatyta efektyviausia konfigūracija kapsuliniai neuroniniai tinklai sprendžiant 3D objektų atpažinimo iš 2D nuotraukų uždaviniui, naudojantis apmokymo laiko ir tikslumo kriterijais.

4. Eksperimentiškai palygintas kapsulinio neuroninio tinklo tikslumas ir apmokymo laikas su konvoliuciniu neuroniniu tinklu, naudojantis apmokymo laiko ir tikslumo kriterijais.

Darbas remiasi šiomis prielaidomis:

1. Kiekvienam 2D paveikslėliui yra priskirta jam jį atitinkantis 3D objektas.
2. Kiekvienas 3D objektas turi bent po vieną jį atitinkantį 2D paveikslėlį.

Šio darbo turinys yra pateiktas taip. Pradedama nuo 3D objektų atpažinimo iš 2D nuotraukų uždavinio aprašymo ir egzistuojančių sprendimų apžvalgos pirmame skyriuje. Tada yra pateikiamas konvoliucinio neuroninio tinklo veikimo aprašymas antrame skyriuje. Toliau trečiame skyriuje yra kapsulinio neuroninio tinklo veikimo aprašymas, išbandytos konfigūracijos ir jų rezultatai. Ketvirtame skyriuje yra aprašomi duomenys naudoti eksperimentams ir patys eksperimentai ir jų rezultatai. Paskutiniame skyriuje yra pateikiamos išvados.

1. 3D objektų atpažinimo iš 2D nuotraukų uždavinys, jo sprendinių pavyzdžiai

3D objektų atpažinimo iš 2D nuotraukų uždavinys - tai klasifikavimo uždavinys, kuriame pateiktos 2D nuotraukos, kuriose yra atvaizduotas 3D objektas iš atsitiktinio apžvalgos taško, turi būti priskirtas 3D modeliui, kuris yra atvaizduotas toje 2D nuotraukoje.

Klasifikavimo uždavinys - tai uždavinys, kuriame kuriamas metodas, kaip nustatyti pavyzdžio, iš tiriamos srities populiacijos, klasę. 3D objektų atpažinimo iš 2D nuotraukų uždavinio atveju, tiriamą sritį yra 2D nuotraukos, kuriose yra atvaizduotas 3D objektas iš bet kurio apžvalgos taško ir klasė - 3D objektas. Taip pat, šio darbo atveju, metodas yra dirbtinio neuroninio tinklo (kapsulinio arba konvoliucinio) apmokytas modelis.

Kaip jau minėta įvade, šiam uždaviniui spręsti efektyviausia yra naudoti mašininio mokymo metodą, kurio mokymo duomenys yra tik 2D nuotraukos, o 3D objektai bus tik duomenų klasės. Darbe [WSK⁺15] atlikto tyrimo rezultatai parodo, kad pateiktas sprendimas, kuriame 3D objektų atpažinimas yra konstruojamas naudojantis tik 2D nuotraukomis, yra tikslesnis 8% (77% → 85%). Šaltinyje [SMK⁺15] yra teigiama, kad to priežastis yra reliatyviai efektyvesnis 2D nuotraukų informacijos saugojimas negu 3D modelių. Todėl, kad, nors 3D modelis turi visą informaciją apie atvaizduotą 3D objektą, tačiau tam, kad panaudoti vokselinę 3D objekto reprezentaciją mašiniame mokyme, kurio mokymas su pakankamai didele duomenų imtimi užtruktų racionalių laiko tarpą, tenka ženkliai sumažinti 3D modelio rezoliuciją. Pavyzdžiui, 3D modelio, kurio rezoliucija yra $30 \times 30 \times 30$ vokseliai, įvesties dydis yra apytiksliai lygus 2D paveikslėlio, kurio rezoliucija yra 164×164 pikseliai. Tad šiuo atveju, 3D modelis yra apdorojamas per tiek pat laiko kaip ir 2D paveikslėlis, bet modelio rezoliucija yra apytiksliai 5.5 karto mažesnė. Todėl mašininio mokymo metodas, kurio mokymo duomenys yra 3D modelis, gauna mažesnės raiškos įvestį, negu metodas, kurio mokymo duomenys yra 2D paveikslėliai.

Vienas seniausių šio uždavinio sprendinių, taikantis tokią metodologiją, yra aprašytas darbe [MN95]. Šis sprendinys atpažįsta 3D objektus lygindamas jų vaizdus, kurie buvo suformuoti iš didelės imties 2D nuotraukų, parametrizuotoje eigenerdvėje (angl. eigenspace). Šios nuotraukos buvo sugeneruotos iš 3D modelių naudojant skirtingus apžvalgos taškus ir apšvietimus. Kitas pavyzdys, kuris yra gana populiarus kompiuterinėje grafikoje, yra šviesos lauko deskriptorius (angl. light field descriptor), kuris yra aprašytas darbe [CTS⁺03]. Šis sprendinys išgauna geometrinius ir Fourier'io deskriptorius iš 3D objektų siluėtų, kurie buvo sugeneruoti iš 3D modelių, naudojant skirtingus apžvalgos taškus. Darbe [MSD⁺02] aprašytas šio uždavinio sprendimas, kuris 3D objekto siluetus išskaido į dalis ir išsaugo juos į orientuotą beciklį grafą (angl. directed acyclic graph), šoko grafą. Kitas pavyzdys aprašytas darbe [CK04], naudoja panašumo metrikas (angl. similarity metrics), kurios yra pagrįstos kreivių palyginimu (angl. curve matching) ir sugrupuotomis panašiomis 2D nuotraukomis.

Šiuo metu, 3D objektų atpažinimo iš 2D nuotraukų uždaviniui spręsti, optimaliausius rezultatus, laiko ir tikslumo atžvilgiu, pasiekęs mašininio mokymo metodu pagrįstas sprendimas yra konvoliuciniai dirbtiniai neuroniniai tinklai. Tai eksperimentu buvo įrodyta darbe [SMK⁺15]. Šia-

me eksperimente buvo palyginti įvairios konvoliucinių neuroninių tinklų tipai sprendžiant šį uždavinį ir geriausią rezultatą pasiekęs tipas buvo daugiavaizdis (angl. multi-view convolutional network) konvoliucinis neuroninis tinklas, kurio tikslumas buvo 90.1%.

2. Konvoliucinių dirbtinių neuroninių tinklų apžvalga

3. Kapsulinių dirbtinių neuroninių tinklų apžvalga ir kalibracija

4. Kapsulinių ir konvoliucinių neuroninių tinklų eksperimentiniai tyrimai

Rezultatai ir išvados

Rezultatų ir išvadų dalyje išdėstomi pagrindiniai darbo rezultatai (kažkas išanalizuota, kažkas sukurta, kažkas įdiegta), pateikiamos išvados (daromi nagrinėtų problemų sprendimo metodų palyginimai, siūlomos rekomendacijos, akcentuojamos naujovės).

Literatūra

- [CK04] C. M. Cyr and B. B. Kimia. A similarity-based aspect-graph approach to 3d object recognition. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.4344&rep=rep1&type=pdf>, 2004. 3820 KB, accessed 2020-08-15.
- [CTS⁺03] D. Chen, X. Tian, Y. Shen, and M. Ouhyoung. On visual similarity based 3d model retrieval. <http://www.cs.jhu.edu/~misha/Papers/Chen03.pdf>, 2003. 1910 KB, accessed 2020-08-15.
- [Mas18] F. Vitor Suzano Massa. Relating images and 3d models with convolutional neural networks. <https://pastel.archives-ouvertes.fr/tel-01762533/document>, 2018. 7.917 KB, accessed 2020-05-23.
- [MC17] R. Mukhometzianov and J. Carrillo. Capsnet comparative performance evaluation for image classification. <https://arxiv.org/ftp/arxiv/papers/1805/1805.11195.pdf>, 2017. 899 KB, accessed 2020-05-23.
- [MN95] H. Murase and S. K. Nayar. Visual learning and recognition of 3-d objects from appearance. <http://murase.m.is.nagoya-u.ac.jp/~murase/pdf/704-pdf.pdf>, 1995. 3041 KB, accessed 2020-08-15.
- [MSD⁺02] D. Macrini, A. Shokoufandeh, S. Dickinson, K. Siddiqi, and S. Zucker. View-based 3-d object recognition using shock graphs. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.852&rep=rep1&type=pdf>, 2002. 155 KB, accessed 2020-08-15.
- [nNH17] S. Sabour nad N. Frosst and G. E. Hinton. Dynamic routing between capsules. <https://arxiv.org/pdf/1710.09829.pdf>, 2017. 899 KB, accessed 2020-05-23.
- [SMK⁺15] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Su_Multi-View_Convolutional_Neural_ICCV_2015_paper.pdf, 2015. 1.439 KB, accessed 2020-05-23.
- [WSK⁺15] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: a deep representation for volumetric shapes. https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Wu_3D_ShapeNets_A_2015_CVPR_paper.pdf, 2015. 2.451 KB, accessed 2020-05-23.

Santrumpos

Sąvokų apibrėžimai ir santrumpų sąrašas sudaromas tada, kai darbo tekste vartojami specialūs terminai, reikalaujantys paaiškinimo, ir rečiau sutinkamos santrumpos.

Priedas Nr. 1

Niauroninio tinklo struktūra

Priedas Nr. 2

Eksperimentinio palyginimo rezultatai

1 lentelė. Lentelės pavyzdys

Algoritmas	\bar{x}	σ^2
Algoritmas A	1.6335	0.5584
Algoritmas B	1.7395	0.5647