

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS
INFORMATIKOS KATEDRA

Generatyviniai besivaržantys neuroniniai tinklai 3D modeliams generuoti iš 2D nuotraukų

Generative Adversarial Networks for 3D Model Reconstruction from 2D Images

Magistro baigiamojo darbo planas

Atliko: Marius Bieliauskas (parašas)

Darbo vadovas: dr. Olga Kurasova (parašas)

Turinys

Santrauka	2
Summary	3
1. Darbo tema ir aktualumas	4
1.1. Tyrimo objektai	4
1.1.1. GAN neuroniniai tinklai	4
1.1.2. Fotogrametrija	5
1.2. Aktualumas	5
2. Tikslas, uždaviniai ir laukiami rezultatai	8
2.1. Darbo tikslas	8
2.2. Uždaviniai	8
2.3. Tyrimo hipotezė ir laukiami rezultatai	8
Literatūra	9

Santrauka

Darbe planuojama įgyvendinti dirbtinio intelekto metodą kuris, naudojant generatyvinius besivaržančius tinklus (angl. Generative adversarial networks arba GAN), galėtų iš vienos ar kelių dvimačių nuotraukų sukonstruoti trimatę fotografuoto objekto reprezentaciją. Viena naujausių dirbtinio intelekto technologijų pritaikoma artimojo nuotolio fotogrametrijos mokslui siekiant išgauti geresnius objektų skaitmeninimo rezultatus nei tradiciniai metodai.

Raktiniai žodžiai: Dirbtinis intelektas, GAN, fotogrametrija, 3D skaitmeninimas.

Summary

The plan for this master's thesis is to implement an artificial intelligence based method which, using generative adversarial networks (GAN), will be able to construct three-dimensional representations of objects captured in one or more photographs. One of the latest technologies in artificial intelligence will be applied to the science of close-range photogrammetry in order to achieve better digitization results than traditional methods.

Keywords: Artificial intelligence, GAN, photogrammetry, 3D digitization.

1. Darbo tema ir aktualumas

Magistro baigiamajam darbui pasirinkta darbo tema "Generatyviniai besivaržantys neuroniniai tinklai 3D modeliams generuoti iš 2D nuotraukų". Šio darbo metu bus nagrinėjami dirbtinio intelekto metodai, konkrečiai generatyviniai besipriešinantys neuroniniai tinklai, norint išgauti trimatę reprezentaciją daiktams, kurie yra vienoje ar keliuose dvimatėse nuotraukose.

1.1. Tyrimo objektai

Darbo metu bus tiriami generatyviniai besivaržantys neuroniniai tinklai ir jų pritaikymas fotogrametrijos mokslui, tiksliau, fotogrametrijos šakai, kuri koncentruojasi į trimatės informacijos išgavimą iš dvimačių duomenų.

1.1.1. GAN neuroniniai tinklai

Generatyviniai besivaržantys neuroniniai tinklai (angl. Generative Adversarial Networks), toliau GAN, buvo pirmą kartą pasiūlyti 2014 metais Ian J. Goodfellow straipsnyje tuo pačiu pavadinimu [GPM⁺14]. Naudojant šį metodą vienu metu yra mokomi du neuronų tinklų modeliai, kurie tarpusavyje konkuruoja. Pirmasis modelis yra generatyvinis ir bando pagal naudotus mokymo duomenis sukurti tokią išvestį, kuri apgautų kitą, diskriminuojantį modelį, kuris gavęs generatyvinio modelio rezultatus, turi atskirti, ar jie priklauso realiems duomenims, ar jie yra sugeneruoti. Taip gaunamas modelių tarpusavio nulinės sumos žaidimas, kuriame vieno modelio sėkmė yra kito modelio nesėkmė.

Dirbtinio intelekto ir tiksliau gilaus mokymo srityje daugiausiai sėkmės yra susilaukę diskriminuojantys (angl. discriminative) modeliai, kurie yra naudojami klasių priskyrimui prie gautų duomenų, tačiau generatyviniai modeliai, kurie gali sukurti naują informaciją iš turimų duomenų ilgą laiką atsiliko savo kokybe. GAN tinklai buvo prasiveržimas generatyvinių modelių srityje, ypač užduotyse susijusiuose su vaizdine informacija, ir privedė prie itin gerų rezultatų kuriant naują informaciją, kuri yra sunkiai atskiriama nuo realios. Vieni labiausiai stebinančių rezultatų gauti žmonių veidų generavimo srityje. 2017 metais Tero Karras ir kt. parašytame straipsnyje [KAL⁺18] gauti fotorealistiški neegzistuojančių žmonių veidai, be jokių įrodymų, kad modelis tiesiog atsiminė ir atkartoją jau egzistuojančius veidus iš mokymo duomenų. Šiuos rezultatus galima pamatyti 1 paveikslėlyje.



1 pav. Straipsnyje [KAL⁺18] gauti rezultatai naudojant CelebA-HQ duomenų rinkinį.

GAN tinklai taip pat gali būti pritaikomi įvairiais kitais būdais, kaip nuotraukų kūrimui iš teksto [ZXL⁺17], nuotraukų koregavimui, kaip, pavyzdžiui, dienos metu kurtos nuotraukos pavertimas į naktinę nuotrauką [IZZ⁺17], taip pat nuotraukų kokybės tobulinimui [LTH⁺17] ir kt.

1.1.2. Fotogrametrija

Fotogrametrija – tai mokslo šaka, tirianti objektų formos, matmenų ir (arba) padėties nustatymą iš fotografinių nuotraukų. Kai kalbama apie fotogrametriją, dažnai kalbama apie tradicinę iš oro arba iš orbitos gautų žemės paviršiaus nuotraukų tyrimą, tačiau šiam darbui pritaikoma artimojo nuotolio fotogrametrija, o tiksliau trimačių modelių išgavimas iš nuotraukų. Vienas standartinių fotogrametrijos metodų pritaikymo pavyzdys matomas Jūratės Sužiedelytės-Visockienės ir kt. 2015 metų straipsnyje [SBM⁺15] kuriame bandoma išgauti trimatę informaciją apie architektūros paveldo deformacijas. Straipsnyje minima, kad nors su lazeriniais skenavimais gaunami geresni trimačio skaitmeninimo rezultatai, dažnai lazerinio skenavimo galimybės nėra ir turi būti atliekama nuotraukų analizė.

1.2. Aktualumas

GAN tinklai yra viena naujausių ir aktualiausių dirbtinio intelekto technologijų, dirbanti su vaizdine medžiaga. Viena mažiau ištyrinėtų sričių susijusių su GAN tinklais yra jų pritaikymas fotogrametrijai ir trimačių modelių kūrimui. 3D modeliavimo sujungimas su GAN technologijomis aprašomas JiaJun Wu ir kt. straipsnyje [WZX⁺16] ir V.V. Kniaz ir kt. straipsnyje [KRK19]. Šis sujungimas gali būti pritaikomas kultūrinio paveldo atkūrimo srityse, pramogų ir ypač kompiuterinių žaidimų srityse, bei kituose srityse, kuriose svarbus objektų 3D skaitmeninimas.

Pirmajame iš minėtų straipsnių [WZX⁺16] labiau susitelkiama į trimačių modelių generavimą pagal turimus esamus modelius, tačiau taip pat kuriamas 3D-VAE-GAN metodas buvo pritaikytas prie ankstesnio 3D skaitmeninimo iš nuotraukų darbo [GFR⁺16] naudojant IKEA baldų duomenų

rinkinį [LPT13]. Pritaikius GAN tinklus, gautas iki 13 % vidutinis patobulėjimas anksčiau geriausiems gautiems rezultatams (žr. 1 lentelė ir 2 pav.).

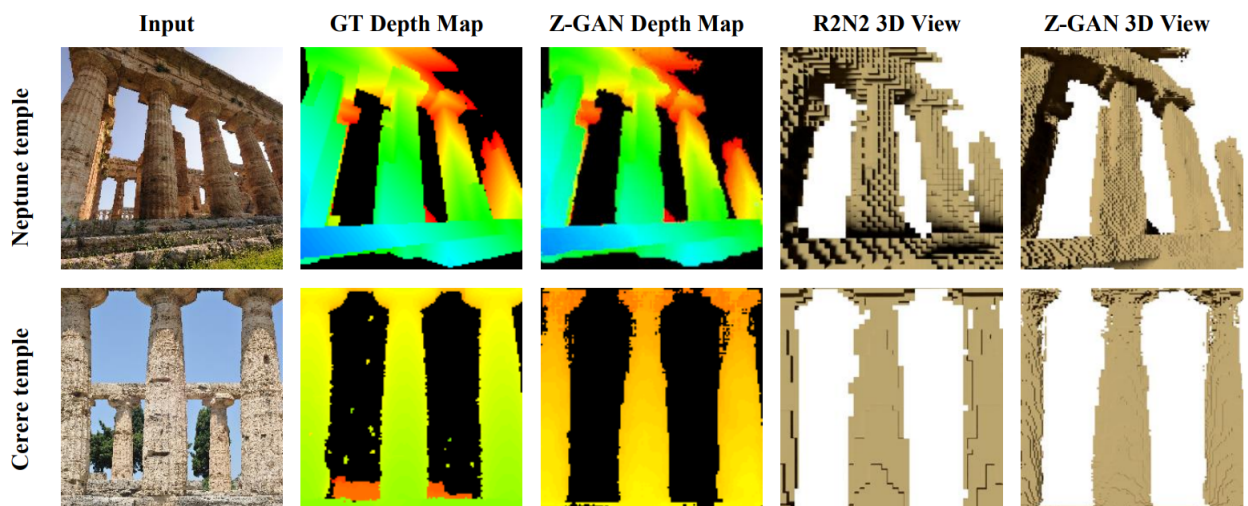
1 lentelė. Wu ir kt. straipsnyje [WZX⁺16] gauti 3D modeliavimo rezultatai (Tikslumas procentais).

Metodas	Lova	Knygų lentyna	Kėdė	Darbo stalas	Sofa	Stalas	Vidurkis
AlexNet-fc8 [GFR ⁺ 16]	29,5	17,3	20,4	19,7	38,8	16,0	23,6
AlexNet-conv4 [GFR ⁺ 16]	38,2	26,6	31,4	26,6	69,3	19,1	35,2
T-L Network [GFR ⁺ 16]	56,3	30,2	32,9	25,8	71,7	23,3	40,0
3D-VAE-GAN (geriausias rezultatas)	63,2	46,3	47,2	40,7	79,8	42,3	53,1



2 pav. Wu ir kt. straipsnyje [WZX⁺16] gauti 3D modeliavimo rezultatai.

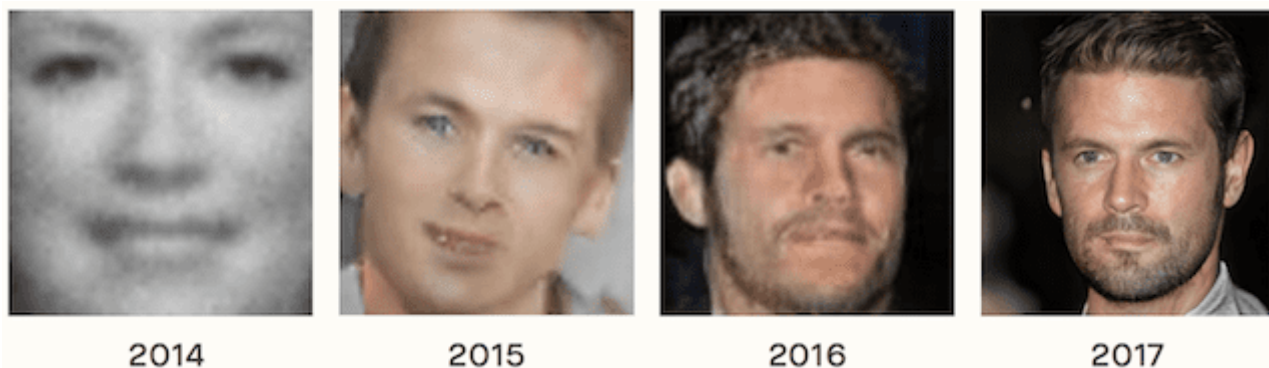
Antrajame straipsnyje [KRK19] remiantis pirmajame ir kituose straipsniuose atliktais darbais buvo sukurtas bendresnis metodas Z-GAN, kuris galėtų kurti modelius anksčiau nematytoms formoms. Tai buvo pritaikyta kultūrinio ir architektūrinio paveldo nuotraukose esančių objektų skaitmeninimui (žr. 3 pav.). Šiuo metodu gauti rezultatai buvo aukštesnės raiškos nei kiti (128x128x128 vietoje 64x64x64) bei buvo įgyvendinta galimybė panaudoti kelias nuotraukas vienam trimačiam modeliui, taip gaunant geresnius rezultatus.



3 pav. Kniaz ir kt. straipsnyje [KRK19] gauti 3D modeliavimo rezultatai.

Jeigu 3D modeliavimo srityje galėtų būti matomas panašus patobulėjimas kaip ir kituose GAN srityse, pvz. veidų kūrimo rezultatuose (žr. 4 pav.), šios technologijos galėtų smarkiai sumažinti dabartinius kaštus visose srityse, kuriose reikalingi daiktų ar aplinkų trimačiai skaitmeninimai,

kadangi šiuo metu geriausios 3D skenavimo technologijos reikalauja brangių gylio matavimo ir kitokių įrankių, kad išgauti reikiamą kiekį informacijos. Atsižvelgiant į tai, šio darbo tema ir tyrimo objektai yra aktualūs tarptautiniu mastu ir potencialiai sugebėjus patobulinti esamus metodus būtų galima prisidėti prie naujausių technologijų vystymo.



4 pav. GAN tinklų tobulėjimas veidų vaizdų kūrimo srityje [BAC⁺18]

2. Tikslas, uždaviniai ir laukiami rezultatai

2.1. Darbo tikslas

Šio darbo tikslas yra įgyvendinti metodą trimačių modelių išgavimui iš dvimačių nuotraukų, naudojant generatyvinius besivaržančius neuroninius tinklus, kuris būtų pranašesnis lyginant su esamais metodais.

2.2. Uždaviniai

Uždaviniai, kurie bus atliekami siekiant darbo tikslo:

1. Įgyvendinti veikiantį GAN tinklą, kuris galėtų iš dvimačių nuotraukų išgauti fotografuojamų objektų trimates reprezentacijas.
2. Optimizuoti įgyvendintus neuroninius tinklus, kad gauti kiek įmanoma geresnius rezultatus tiksliai su viena nuotrauka.
3. Pridėti galimybę naudoti kelias nuotraukas vieno objekto modelio sukūrimui, siekiant, kad panaudojus mažiau nuotraukų būtų gaunami geresni rezultatai nei su tradiciniais daugelio nuotraukų skaitmeninimo metodais.
4. Pritaikyti automatinio išlyginimo (angl. smoothing) algoritmą, kuris galėtų patobulinti gautų 3D modelių raišką ir kokybę.

2.3. Tyrimo hipotezė ir laukiami rezultatai

Hipotezė: naudojant generatyvinius besivaržančius neuroninius tinklus galima patobulinti fotogrametrijos metodus trimačių modelių skaitmenizavimui iš dvimačių nuotraukų.

Laukiami rezultatai: Metodai naudojantys generatyvinius besivaržančius neuroninius tinklus, kad skaitmeninti vienoje ar keliuose nuotraukose esančius objektus į trimatę erdvę, bei 3D modelių išlyginimo metodas. Taip pat programinė įranga, kuri šiuos metodus sujungtų ir gautų fotografuojamų objektų trimates reprezentacijas reikalaujančias kuo mažiau rankinio koregavimo, kad atitiktų tikrovę.

Literatūra

- [BAC⁺18] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner ir k.t. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *ArXiv*, abs/1802.07228, 2018. URL: <https://arxiv.org/abs/1802.07228>.
- [GFR⁺16] Rohit Girdhar, David F. Fouhey, Mikel D. Rodriguez ir Abhinav Kumar Gupta. Learning a Predictable and Generative Vector Representation for Objects. *ArXiv*, abs/1603.08637, 2016. URL: <https://arxiv.org/abs/1603.08637>.
- [GPM⁺14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville ir Yoshua Bengio. Generative Adversarial Nets. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence ir K. Q. Weinberger, redaktoriai, *Advances in Neural Information Processing Systems*, tom. 27. Curran Associates, Inc., 2014. URL: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.
- [IZZ⁺17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou ir Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p.p. 5967–5976, 2017. DOI: 10.1109/CVPR.2017.632. URL: <https://arxiv.org/abs/1611.07004>.
- [KAL⁺18] Tero Karras, Timo Aila, Samuli Laine ir Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *International Conference on Learning Representations*, 2018. URL: <https://openreview.net/forum?id=Hk99zCeAb>.
- [KRK19] Vladimir Kniaz, Fabio Remondino ir V. Knyaz. GENERATIVE ADVERSARIAL NETWORKS FOR SINGLE PHOTO 3D RECONSTRUCTION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9:403–408, 2019-01. DOI: 10.5194/isprs-archives-XLII-2-W9-403-2019. URL: https://www.researchgate.net/publication/330766571_GENERATIVE_ADVERSARIAL_NETWORKS_FOR_SINGLE_PHOTO_3D_RECONSTRUCTION.
- [LPT13] Joseph J. Lim, Hamed Pirsiavash ir Antonio Torralba. Parsing IKEA Objects: Fine Pose Estimation. *2013 IEEE International Conference on Computer Vision*, p.p. 2992–2999, 2013. DOI: 10.1109/ICCV.2013.372. URL: https://people.csail.mit.edu/torralba/publications/ikea_iccv2013.pdf.
- [LTH⁺17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang ir Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*:105–114, 2017. URL: <https://arxiv.org/abs/1609.04802>.

- [SBM⁺15] Jūratė Sužiedelytė-Visockienė, Renata Bagdžiūnaitė, Naglis Malys ir Vida Maliene. Close-range photogrammetry enables documentation of environment-induced deformation of architectural heritage. *Environmental Engineering and Management Journal*, 14, 6, 2015. URL: <http://eprints.nottingham.ac.uk/id/eprint/34653>. Full text of article not available online. No doi. Permission given by journal publishers to deposit in repository. KJB 11.07.2016.
- [WZX⁺16] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman ir Joshua B. Tenenbaum. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *NIPS*, 2016. URL: <https://arxiv.org/abs/1610.07584>.
- [ZXL⁺17] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang ir Dimitris N. Metaxas. StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*:5908–5916, 2017. URL: <https://arxiv.org/abs/1612.03242>.