

Generatyviniais besivaržančiais tinklais sukurtų 3D modelių tikslumo ir tikroviškumo įvertinimas

Justina Maslovaite

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Naugarduko g. 24, LT-03225 Vilnius
justina.maslovaite@mif.stud.vu.lt

Santrauka. Šiame straipsnyje pateikiami trimačių modelių generavimo GAN neuroniniu tinklu įvertinimo metodai. Dažnu atveju generatyviniai tinklai neturi aiškių gairių ir patarimų kaip įvertinti generuojamus modelius. Toks trūkumas apsunkina vertinimą ir reikalauja remtis vizualiu kiekvieno modelio patikrinimu. Norint iš tikrųjų įvertinti 3D modelio kokybę, modelio vertinimą reikia traktuoti kaip pastovų procesą ir apibrėžti nekintamus kriterijus. Šių kriterijų pagalba visiems tiriamiems modeliams būtų sukurtos vienodos vertinimo sąlygos, o rezultatai nepriklausytų nuo vertinančio žmogaus nešališkumo ir patirties. GAN sugeneruoto rezultato tikslumui ir realumui įvertinti pritaikomos Sørensen – Dice ir Jaccard atstumo metrikos. Iškeliami esminiai šių metrikų taikymo iššūkiai ir nagrinėjami veiksniai darantys įtaką metrikų galutiniam įverčiui. Gauti rezultatai sudaro prielaidas tolimesnių tyrimų vykdymui.

Raktiniai žodžiai: kompiuterinė rega, gilus mokymasis, 3D modelio generavimas, 3D modelio įvertinimas.

1 Įvadas

Dirbtinio intelekto ir gilaus mokymosi pritaikymas kasdienybėje įgauna pagreitį ir sulaukia vis daugiau susidomėjimo, o kompiuterio grafikai tobulėjant, auga tikslesnės virtualios realybės poreikis. Goodfellow ir kt. [1] 2014 metais pristatytas GAN (angl. Generative adversarial network) tinklas atvėrė galimybes tikroviškų paveikslėlių generavimui ir tapo vienu sėkmingiausių generavimo modelių per pastarąjį dešimtmetį. Lengvesnis 3D duomenų prieinamumas, pažanga 3D sensorinės technologijos srityje ir toks gilaus mokymosi pasisėkimas paskatino kompiuterinės grafikos bendruomenę gilintis į neuroninių tinklų trimatėje erdvėje realizaciją. Tačiau tyrimų kiekis atliekant užduotis tokias kaip vaizdų generavimas, klasifikavimas ar objektų aptikimas dvimatėje erdvėje gerokai pranoksta tyrimų kiekį minėtas užduo-

tis atlikti trimatėje erdvėje. Sudėtinga 3D duomenų prigimtis ir struktūriniai skirtumai priverčia ieškoti kūrybingų informacijos apdorojimo sprendimų. Esant gausiam programinės įrangos ir trimačių duomenų išsaugojimo būdų pasirinkimui, iškyla klausimas, kaip įvertinti 3D modelių kokybę ir kaip šiuos modelius tarpusavyje palyginti.

Šio tyrimo tikslas – palyginti trimačių GAN tinklo sugeneruotų modelių kokybės ir panašumo vertinimo kriterijus. Tyrimo metu analizuojamos dažniausiai naudojamos trimačių modelių reprezentacijos formos ir duomenų paruošimo metodai, pritaikomas GAN neuroninis tinklas trimačių modelių generavimui.

2 Susiję darbai

3D GAN sugeneruotų modelių realumo ir tikslumo įvertinimas yra nelenava užduotis, kuri reikalauja kiekybiškai palyginti sugeneruotų modelių ir realaus pasaulio objektų, kuriuos jie turi reprezentuoti, panašumą. Dažnu atveju toks vertinimas neapsieina be vizualaus tokių modelių tikrinimo ir yra paveikiamas vertintojo nuomonės. Norint šį procesą automatizuoti, reikalingos papildomos metrikos, kurių pagalba vertinimo procesas būtų nepriklausomas, o vertinimas būtų pagrįstas realaus pasaulio objektų ir 3D modelių panašumu. Šios metrikos naudoja algoritmus, kad palygintų sugeneruotų modelių vaizdines savybes su realaus pasaulio objektų ypatybėmis priskiriant balą, atspindintį jų panašumą.

Tim Salimans ir kt. [2] 2016 metais siekdami automatizuoti daugybės GAN sugeneruotų vaizdų palyginimo procesą pasiūlė IS (Inception distance) metriką. IS matuoja sugeneruotų paveikslėlių kokybę remiantis, kaip gerai sukurti egzemplioriai atspindi įvairias duomenų kategorijas ir kaip sugeneruotą egzempliorių įvertino tuo metu geriausias klasifikavimo modelis *Inception v3* [3]. Modelis skiria užtikrintumo balą, nusakantį priklausomybės objekto klasei tikimybę. Šį įvertinimo metodą 2017 metais patobulino Heusel ir kt. [4] pristatydami FID (Fréchet Inception distance). FID paprastai laikomas patikimesniu nei IS, kadangi FID pranašumas – didesnis atsparumas vaizdo raiškos, stiliaus ir turinio pokyčiams, tad šis įvertis gali būti pritaikomas įvairesnėms generavimo užduotims įvertinti. Taip pat FID yra mažiau jautrus duomenų rinkinio dydžiui bei klasių skaičiui, tuo tarpu šie veiksniai gali daryti įtaką IS įverčiui. Nors tiek IS tiek FID sėkmingai pritaikomi generatoriaus kokybei įvertinti, Chong ir kt. [5] pastebi, kad FID pilnai neatspindi klasių įvairovės, ko pasekoje generatoriaus kokybės įvertis gali būti netiks-

lus. Šiuo metu egzistuoja daugiau IS ir FID variacijų, keletas tokių pavyzdžių: KID (Kernel Inception distance), švarus FID (Clean FID), CFID (Conditional FID) ir t.t. IS ir FID automatizuotas paveikslėlių kokybės vertinimo metodas dažnai sutapdavo su vizualiu žmogaus paveikslėlių vertinimu ir galėtų būti alternatyva vizualiam tikrinimui pakeisti.

Dauguma GAN kokybės nustatymo metodų turi bendrą trūkumą – vertinimo gilumas būna nepakankamas ir atskiri nesėkmių atvejai lieka nepastebėti. Sajjadi ir kt. [6] pabrėžia šią problemą paprastu GAN skaitmenų generavimo pavyzdžiu. Paveikslėlyje (1 pav.) matomi GAN generatoriaus sugeneruoti skaitmenys.



1 pav. Sajjadi ir kt. [6] darbe pateiktas generatoriaus nepakankamumo pavyzdys.

Nors FID įvertis tiek kairėje tiek dešinėje yra panašus, kairiajame paveikslėlyje skaitmenų imtis, o dešiniajame skaitmenų kokybė yra riboti. Generatoriaus sukurti skaitmenys atrodo tikslūs ir kokybiški, tačiau šių skaitmenų imtis svyruoja tarp 0 ir 3, o didesnių sugeneruotų skaitmenų atvejų visiškai nematyti. Tuo tarpu dešinėje pusėje skaitmenų įvairovė yra didesnė, tačiau prasta tokių skaitmenų kokybė. Šiai problemai spręsti Sajjadi ir kt. [6] pasiūlyta *Alfa precision and recall* metrika, kuri atsižvelgia į sugeneruotų pavyzdžių kintamumą ir pateikia išsamesnį GAN kokybės įvertinimą. *Alfa precision and recall* parodo ar GAN sukurti vaizdai vidutiniškai yra panašūs į tikrus vaizdus ir ar juose atsispindi apmokymo duomenų įvairovė. Šiuo įverčiu siekiama išspręsti įvairovės trūkumo problemą įvedant derivamą parametą *alfa* tikslumą, kuris kontroliuoja tikrojo ir sugeneruoto egzempliorių sutapimo dydį. *Alfa* tikslumu matuojama sukurtų vaizdų, kurie yra tikroviški, dalis, tuo tarpu *recall* dydžiu matuojamas generatoriaus apmokymo duomenų įvairovės laipsnis. Tačiau šie abu aspektai dažnai vienas kitą veikia, ko pasekoje padidinus modelio generuojamų egzempliorių tikslumą, sumažėja generuojamų pavyzdžių įvairovė ir atvirkščiai.

Belongie ir kt. [7] pristatė Chamfer atstumo įvertį, sukurtą dviejų 3D erdvės taškų rinkinių panašumui nustatyti. *Chamfer* atstumas dažnai naudojamas kompiuterinės regos ir grafikos uždaviniuose, kuomet norima aptikti formą išreikštą taškų rinkiniu. *Chamfer* atstumas apibrėžiamas kaip vidutinis atstumas nuo kiekvieno A rinkinio taško iki artimiausio kaimyno rinkinyje B ir atvirkščiai. Intuityviai metodas nusako kaip arti taškų rinkinių aibės viena nuo kitos. GAN sukurtų egzempliorių kontekste *Chamfer* atstumas gali būti naudojamas siekiant įvertinti sugeneruoto modelio ir tikrojo modelio panašumą. *Chamfer* atstumas atveria galimybes vertinimo 3D erdvėje realizacijai ir yra atsparus triukšmui taškų debesyje. Bendrai ši metrika tapo gana plačiai taikoma 3D formų paieškos uždaviniuose.

3 Trimatis GAN modelis

Tyrimo ir praktinio realizavimo metu trimačių modelių generavimui ir tolimesnei rezultatų analizei pasiremta 3D GAN [8] neuroninio tinklo realizacija atlikta remiantis *PyTorch* biblioteka. Modelio apmokymui ir testavimui, skirtingai nuo originalaus 3D GAN, kuriame naudotas *ShapeNets* [9] duomenų rinkinys, pritaikytas *ModelNet* [10] kėdžių modelių rinkinys. Šio duomenų rinkinio kiekvienas kėdės objektas patalpintas $30 \times 30 \times 30$ kubo dydžio tinklelyje, o modelio testavimo metu generuojamas tikimybių vektorius pasiskirstymas taip pat atitinka $30 \times 30 \times 30$ kubo dydžio tinklelio rezoliuciją. Svarbu paminėti, kad 3D GAN modelio architektūroje naudojamas $64 \times 64 \times 64$ dydžio kubas, tad gaunami rezultatai gali nesutapti su rezultatais pateiktais autorių darbe. Neuroninio tinklo aktyvacijos funkcijos tokios kaip *Adam* ir *Leaky Relu* taip pat ir partijos dydis 32 išlieka tokie patys kaip ir originaliame 3D GAN neuroninio tinklo modelyje. Neuroninio tinklo apmokymo procesą sudarė 500 iteracijų, kur kiekvienos iteracijos pabaigoje išsaugoma modelio būseną, taip pat iteracijos pabaigoje neuroninio tinklo modelis testuojamas ir atvaizduojami šeši sugeneruotų kėdžių pavyzdžiai.

4 Sørensen – Dice ir Jaccard koeficientų pritaikymas

Sørensen – Dice koeficientas [11] ir Jaccard indeksas [12] plačiai naudojami įvairių algoritmų tikslumo nustatymui. Keletas tokių algoritmų pavyzdžių: trimačių objektų 2D vaizduose ar vaizdo įrašuose aptikimas, dviejų tekstinių duomenų rinkinių panašumo nustatymas, mašininio mokymosi modelių tikslumo matavimas, kur lyginamos modelių prognozės su nustatyta tiesa.

Matematiškai Dice panašumo koeficientas tarp dviejų aibių X ir Y išreiškiamas formule:

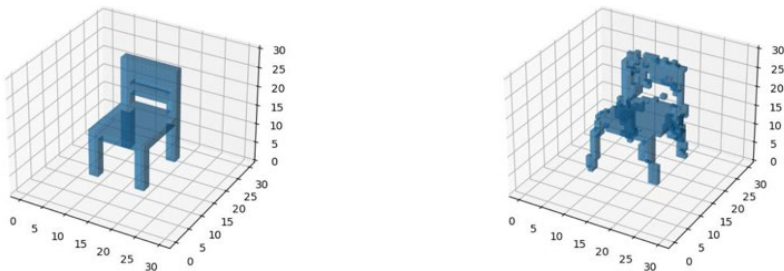
$$Dice(X, Y) = \frac{2 |X \cap Y|}{|X| + |Y|}$$

Dviejų vaizdų palyginime Dice apskaičiuojamas vaizdų bendrų pikselių skaičių dauginant iš dviejų ir dalijant iš abiejų paveikslėlių pikselių skaičiaus sumos. Tuo tarpu Jaccard indeksas apskaičiuojamas kaip dviejų vaizdų bendrų pikselių skaičiaus ir abiejų vaizdų visų pikselių sąjungos santykis. Matematiškai Jaccard indeksas tarp dviejų aibių X ir Y išreiškiamas formule:

$$Jaccard(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

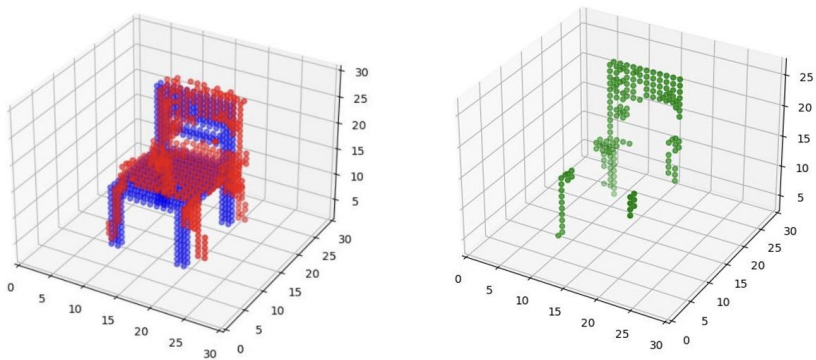
Abi įverčių reikšmės svyruoja tarp 0 ir 1, kur 1 nusako dviejų aibių identišumą. Prieš pasirenkant kažkurį vieną iš šių įverčių, svarbu nuspręsti, kokius ryšius tarp lyginamų rinkinių bandoma užfiksuoti, ar visais atvejais yra apytiksliai toks pat elementų skaičius aibėse, ar kiekvienu iš atvejų rinkinio elementų skaičius ženkliai skirtingas. Jaccard indeksas yra jautresnis elementų skaičiaus skirtumams tarp rinkinių nei Dice koeficientas, net jei vienas iš rinkinių galėtų būti kito poaibis.

Dice ir Jaccard įverčių apskaičiavimui naudojamas modelis iš ModelNet apmokymo duomenų aibės ir GAN sugeneruotas netvarkingas kėdės modelis. Šiame procese abu įverčiai pritaikomi tiesiogiai palyginant binarinius abiejų modelių duomenis naudojantis aukščiau nurodytomis Dice ir Jaccard formulėmis.



2 pav. Originalios kėdės iš *ModelNet* apmokymo duomenų rinkinio ir GAN sugeneruotos kėdės modeliai.

Vizualaus modelių palyginimo metu vertinant 2 pav. atvaizduotus modelius galima būtų spėti, kad kėdžių modeliai ganėtinai panašūs tarpusavyje, tačiau gauti Dice ir Jaccard koeficientų rezultatai rodo, kad šių modelių panašumas siekia tik 0,27 balo Dice atveju ir 0,15 Jaccard indekso atveju. Žemi panašumo įverčiai gali būti paaiškinami vizualiai matoma abiejų modelių sankirta (3 pav.), kurioje galima pastebėti, kad sugeneruotas modelis patalpintas aukščiau originalaus kėdės modelio ir besikertančių taškų dalis yra sąlyginai maža.



3 pav. Abiejų kėdžių modeliai atvaizduoti 3D vokselių tinklelyje. Raudona spalva žymi mas GAN sugeneruotas kėdės modelis, mėlyna – modelis iš *ModelNet* apmokymo duomenų aibės, žalia – modelių bendri taškai.

Kėdžių objektų tarpusavio palyginimo metu išryškėja keletas iššūkių, tokių kaip: kėdžių formos ir stiliaus nepastovumas duomenų apmokymo ir generatoriaus generuojamų pavyzdžių rinkiniuose, taip pat skirtingas objektų dydis bei pozicija vokselių tinklelyje. Aukščiau esančio pavyzdžio atveju (3 pav.) nors ir galima būtų teigti, kad bendras objektų skeletas yra labai panašus, tačiau netinkama sugeneruoto modelio pozicija vokselių tinklelyje ženkliai paveikia abiejų įverčių balus ir nurodo didesnę šių modelių skirtumą.

Vokselių skaičius modelių panašumo nustatymo metu gali daryti reikšmingą įtaką Dice ir Jaccard įverčiams. Esant skirtingam vokselių skaičiui paveikiami abiejų aibių sankirtos ir sąjungos rinkinių dydžiai, o tai savo ruožtu paveikia panašumo koeficientus. Vieno iš kėdės modelių vokselių skaičius atsitiktiniu būdu sumažinamas 15 %, 30 % ir 60 % nuo pradinio modelio dydžio, toks modelis toliau lyginamas su tuo pačiu, pilnu kėdės modeliu.

Pateikti rezultatai (1 lentelė) parodo, kad vokselių skaičiaus skirtumas abiejuose modeliuose daro didelę įtaką panašumo įverčiams ir matomas ryškus Dice koeficiento sumažėjimas mažėjant vokselių skaičiui viename iš modelių. Kaip buvo tikėtasi, Jaccard indeksas šiuo atveju jautriau reaguoja į vokselių skaičiaus pokyčius.

1 lentelė. Vokselių skaičiaus įtaka Dice ir Jaccard įverčiams.

Vokselių skaičius modelyje	Dice įvertis	Jaccard įvertis
720 (100 %)	1,00	1,00
615 (85 %)	0,90	0,85
505 (70 %)	0,82	0,69
435 (60 %)	0,53	0,41

Vertinamų modelių posūkio laipsnis yra dar vienas iš veiksnių darančių įtaką galutiniam Dice ir Jaccard įverčių balui. Dirbant su trimatėmis formomis, dažnu atveju tenka įvertinti ne tik šių formų poziciją, tačiau ir posūkio laipsnį trimačiame vokselių tinklelyje. Formos tarpusavyje gali būti identiškios, tačiau skirtingas šių formų posūkio laipsnis beveik visais atvejais sąlygoja mažesnius koeficientų įverčius. Posūkio vertinimo metu modelis lyginamas su pačiu savimi atliekant vieno iš lyginamų modelių rotaciją vis didesniu kampų, bendrai modelis pasukamas 5 kartus nuo pradinės modelio pozicijos. Galima matyti (2 lentelė), kad kiekvienu atveju Dice ir Jaccard įverčių reikšmės yra veikiamos posūkio net esant visiškam modelio identiškumui.

2 lentelė. Posūkio įtaka Dice ir Jaccard įverčiams.

Modelio posūkio laipsnis	Dice įvertis	Jaccard įvertis
0°	1,00	1,00
30°	0,50	0,33
90°	0,47	0,31
180°	0,69	0,52
210°	0,41	0,26
270°	0,47	0,31

5 Išvados

Sudėtinga 3D formų prigimtis ir bendrų vertinimo kriterijų trūkumas apsunkina tokių formų taikymą generavimo kontekste, tad dauguma vertinimo metodų remiasi mechanišku rezultatų tikrinimu ir priklauso nuo kitų klasifikavimo modelių. Tyrimo metu apžvelgti dažniausiai naudojami paveikslėlių ir trimačių modelių panašumo nustatymo metodai. Trimačių vokselių formų palyginimui pritaikyti Dice ir Jaccard panašumo koeficientai. Iškeltos esminės problemos susijusios su tokių formų palyginimu kaip: formų stiliaus įvairumas, dydžio, pozicijos ir rotacijos skirtumai veikiantys tiek Dice tiek Jaccard panašumo įverčius. Norint pilnai automatizuoti GAN sugeneruotų trimačių formų tikroviškumo ir realumo vertinimo procesą, Dice ir Jaccard panašumo koeficientai galėtų būti vienas iš sprendimų, tačiau svarbus pirminis tokių formų normalizavimas, kuris galėtų būti tolimesnė šio tyrimo dalis.

Literatūra

- [1] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville ir Yoshua Bengio (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.
- [2] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford ir Xi Chen (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.
- [3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke ir Andrew Rabinovich (2014). Going Deeper with Convolutions. CoRR, abs/1409.4842. arXiv: 1409. 4842. URL: <http://arxiv.org/abs/1409.4842>.
- [4] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler ir Sepp Hochreiter(2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- [5] Min Jin Chong ir David Forsyth (2020). Effectively Unbiased FID and Inception Score and Where to Find Them. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020-06.
- [6] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet ir Sylvain Gelly (2018). Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31.
- [7] S. Belongie, J. Malik ir J. Puzicha (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509-522, 2002. DOI: 10.1109/34.993558.
- [8] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T Freeman ir Joshua B Tenenbaum (2016). Learning a probabilistic latent space of object shapes via 3d generative adversarial modeling. *Advances in Neural Information Processing Systems*, p.p. 82-90.

- [9] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang ir Jianxiong Xiao (2015). 3d shapenets: A deep representation for volumetric shapes. Proceedings of the IEEE conference on computer vision and pattern recognition, p.p. 1912–1920.
- [10] Kashi Venkatesh Vishwanath, Diwaker Gupta, Amin Vahdat ir Ken Yocum (2009). Modelnet: Towards a datacenter emulation environment. 2009 IEEE Ninth International Conference on Peer-to-Peer Computing, p.p. 81–82. IEEE.
- [11] Alonso Gragera ir Vorapong Suppakitpaisarn (2016). Semimetric properties of sørensen-dice and tversky indexes. WALCOM: Algorithms and Computation: 10th International Workshop, WALCOM, Kathmandu, Nepal, March 29–31, 2016, Proceedings 10, p.p. 339–350. Springer.
- [12] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn ir Supachanun Wanapu (2013). Using of Jaccard coefficient for keywords similarity. Proceedings of the international multiconference of engineers and computer scientists, tom. 1 numeris 6, p.p. 380–384.