

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
INFORMATIKOS INSTITUTAS
PROGRAMŲ SISTEMŲ STUDIJŲ PROGRAMA

**Vaizdo žaidimo eismo vaizdų generavimas
naudojant nesuporuotų vaizdų transformacijas**

**Video game traffic image generation using unpaired image
to image translation**

Bakalauro baigiamojo darbo planas

Atliko:	Gytis Oksas	(parašas)
Darbo vadovas:	j. asist. Boleslovas Dapkūnas	(parašas)
Darbo recenzentas:	doc. dr. Linas Petkevičius	(parašas)

Vilnius – 2023

Įvadas

Vaizdų transformacija (angl. *image-to-image translation*) yra plačiai taikoma metodika kompiuterių regos ir paveikslų apdorojimo problemoms spręsti [PLQ⁺21]. Metodai, sprendžiantys šias problemas, gali būti taikomi parinktų specifinių paveikslėlių požymių keitimui, trūkstamų paveikslėlio pikselių užpildymui arba realistinių vaizdų generavimui. Ši taikymo sritis gali būti naudinga kuriant žaidimų meninį turinį (pavyzdžiui, tekstūras, realistinių objektų dizaino atitikmenis žaidimui), kuriant animacijas. Ir nors realių vaizdų transformavimas į nerealistinių vaizdų domeną nėra naujas (pavyzdžiui, veido nuotraukas transformuoti į animacinių filmukų stilių, kaip yra taikoma MSPC modelio pavyzdžiuose [XXW⁺22]), jis vis dar yra aktyviai analizuojamas, kadangi yra likusių ir pastoviai išskylančių problemų ir iššūkių (pavyzdžiui, vaizde daug triukšmo, besikartojantys artefaktai, nelogiški piešiniai). Vaizdo žaidimų vaizdų transformavimas į realistiškus vaizdus yra spęstas, tačiau atvirkštinis variantas yra gan mažai išnagrinėtas.



1 pav. Vaizdai prieš ir po nuotraukos transformavimo, atitinkamai.[RAK21]

Tyrimas kuris atkreipė daug dirbtinio intelekto ir vaizdo žaidimų mėgėjų dėmesį yra aprašytas Intel dirbtinio intelekto mokslininkų straipsnyje „Enhancing Photorealism Enhancement“ [RAK21], jame pristatomas metodas, kaip išnaudojus modernias dirbtinio intelekto technologijas yra sukuriamas nuotraukų transformatoriaus modelis, sugebantis žaidimo GTA V automobilių eismo vaizdus paversti foto realistiniais, lyg įrašytais automobilio registratoriumi. Jis vienas pirmųjų įtikinamai, švariai, be prarastų esminių nuotraukų detalių ir be pridėtinių artefaktų sugebėjo transformuoti nuotrauką į realistiškai atrodantį vaizdą (žr. 1 pav.). Tai davė idėją šiam tyrimui –

apsukti puses ir sukurti modelį, kuris pagal tuos pačius atributus sugebėtų transformuoti realybės automobilių eismo vaizdus į vaizdo žaidimo automobilių eismo vaizdus.

Tyrimo metodai ir laukiami rezultatai

Kursiniame darbe buvo panaudoti, jau tapę tradiciniais, generatyviniai adversariniai tinklai[GPM⁺14] (angl. *generative adversarial network*), kurie turėjo savo trūkumų su turinio dingimu po nuotraukos transformacijos (daugiausiai iš nuotraukų buvo ištrinti automobiliai), todėl svarbi tyrimo dalis bus sukurti metodą, leidžiantį atskirti bendrą turinį ir dingstančias klases iš vaizdo (naudojant segmentacinius modelius) ir taip atskirai transformuoti bendrą vaizdą ir dingstantį turinį (t.y. automobilius), ir šiuos vaizdus algoritmiškai sujungti. Šio metodo norimas rezultatas yra kokybiškas foninių detalių transformavimas su išsaugomais automobiliais nuotraukose, t.y. kelių modelių programa, kuri sugeba išsaugoti daugiau detalių nei pavienis MSPC modelis[XXW⁺22] (apmokytas kursinio darbo metu) be papildomos logikos ar skaičiavimų.

Pastaraisiais metais taip pat pradėjo populiarėti difuziniai nuotraukų generavimo modeliai[YZS⁺22]. Jie sugeba kurti nuotraukas su kur kas mažiau triukšmo ir įtikinamesniu turiniu, tačiau šių modelių mokymas ir naudojimas trunka kur kas ilgiau. Dėl šių priežasčių, dalis darbo gali būti skiriama bandymams, ar šie nauji modeliai gali būti pritaikyti ir gali būti naudingesni nei įprasti GAN modeliai, šios užduoties scenarijuje. Norimas rezultatas iš šio eksperimento būtų kokybiškiau arba švariau (mažiau artefaktų, nuotraukos triukšmo) transformuotas vaizdas.

Darbo atlikimo procesas

1. Sukurti vaizdo žaidimo automobilių duomenų rinkinį.
2. Apmokyti MSPC modelį transformuoti tik automobilių nuotraukas
3. Kursiniame darbe apmokytą bendrą MSPC modelį ir MSPC automobilių nuotraukų transformavimo modelį apjungti programoje, kuri naudojanti segmentacinį modelį atskirtų automobilius iš nuotraukos, transformuotų abi nuotraukas ir įklijuotų transformuotą automobilio nuotrauką atgal į bendrą nuotrauką.
4. Rasti ar yra difuzijos modelių sugebančių transformuoti nuotraukas tarp dviejų domenų naudojant nesuporuotų nuotraukų transformaciją, jei rasti pavyktų, apmokyti difuzinį modelį su turimu duomenų rinkiniu.
5. Palyginti ar naujų modelių rezultatai yra geresni negu kursiniame darbe gauto modelio (atsižvelgiant į naujesnių metodų kompleksiskumą ir didesnes sąnaudas).

Darbo tikslas

Šio tyrimo tikslas ir yra sukurti metodą, gebantį transformuoti realias eismo nuotraukas į vaizdo žaidimo vaizdus, tačiau neprarandant esminių semantinių objektų (šiuo atveju – automobilių). Pasirinkta buvo transformuoti nuotraukas į žaidimo „Grand Theft Auto: Vice City“ (toliau trumpinama „GTA:VC“) domeną dėl jo išskirtinio meninio stiliaus ir ryškaus domeno požymių skirtumo lyginant su realaus eismo nuotraukomis. Tyrimo uždaviniai:

1. Sukurti vaizdo žaidimo duomenų rinkinį.

2. Sukurti vaizdo žaidimo automobilių duomenų rinkinį.
3. Sukurti metodą sugebantį atskirai transformuoti tam tikrą nuotraukos semantinę klasę ir gautą rezultatą atgal įklijuoti į nuotrauką.
4. Palyginti modelius.

- [COR⁺16] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, Bernt Schiele. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. arXiv, 2016. Pasiiekiamas per DOI: 10.48550/ARXIV.1604.01685.
- [DBX⁺22] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, Qi Tian. *CenterNet++ for Object Detection*. arXiv, 2022. Pasiiekiamas per DOI: 10.48550/ARXIV.2204.08394.
- [GPM⁺14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. *Generative Adversarial Networks*. arXiv, 2014. Pasiiekiamas per DOI: 10.48550/ARXIV.1406.2661.
- [HRU⁺17] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. 2017. Pasiiekiamas per DOI: 10.48550/ARXIV.1706.08500.
- [YCW⁺18] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, Trevor Darrell. *BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning*. arXiv, 2018. Pasiiekiamas per DOI: 10.48550/ARXIV.1805.04687.
- [YZS⁺22] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong ir kt. *Diffusion Models: A Comprehensive Survey of Methods and Applications*. arXiv, 2022. Pasiiekiamas per DOI: 10.48550/ARXIV.2209.00796.
- [LMB⁺14] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev ir kt. Microsoft COCO: Common Objects in Context. *CoRR*. 2014, t. abs/1405.0312. Pasiiekiamas per arXiv: 1405.0312.
- [PLQ⁺21] Yingxue Pang, Jianxin Lin, Tao Qin, Zhibo Chen. *Image-to-Image Translation: Methods and Applications*. arXiv, 2021. Pasiiekiamas per DOI: 10.48550/ARXIV.2101.08629.
- [RAK21] Stephan R. Richter, Hassan Abu AlHaija, Vladlen Koltun. Enhancing Photorealism Enhancement. *arXiv:2105.04619*. 2021.
- [XXW⁺22] Yanwu Xu, Shaoan Xie, Wenhao Wu, Kun Zhang, Mingming Gong, Kayhan Batmanghelich. Maximum Spatial Perturbation Consistency for Unpaired Image-to-Image Translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022-06, p. 18311–18320.