

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

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

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

Kursinis darbas

Atliko: 4 kurso 5 grupės studentas

Gytis Oksas

(parašas)

Darbo vadovas: j. asist. Boleslovas Dapkūnas

(parašas)

## TURINYS

IVADAS .....	2
1. GENERATYVINIAI ADVERSARINIAI TINKLAI .....	4
1.1. GAN modelis .....	4
1.2. CycleGAN modelis .....	4
1.3. CUT modelis .....	4
1.4. MSPC modelis .....	5
2. EKSPERIMENTAS .....	6
2.1. Duomenų rinkiniai .....	6
2.1.1. Realių vaizdų duomenų rinkinys .....	6
2.1.2. Grand Theft Auto: Vice City duomenų rinkinys .....	6
2.2. Mokymo platformos .....	8
2.3. Mokymo procesas .....	8
2.4. Modelių vertinimo metrika .....	9
2.5. „Dingstančių automobilių“ problema .....	9
2.6. Paros meto keitimosi problema .....	11
2.7. Gauti FID įverčiai .....	11
REZULTATAI IR IŠVADOS .....	14
SANTRUMPOS .....	15
LITERATŪRA .....	16
PRIEDAI .....	16
1 priedas. Darbo saugykla .....	17
2 priedas. Nuotraukų pavyzdžiai (pirmosios duomenų rinkinio iteracijos) .....	18
3 priedas. Nuotraukų pavyzdžiai (antrosios duomenų rinkinio iteracijos) .....	19

# Ivadas

Vaizdų transformacija (angl. *image-to-image translation*) yra plačiai taikoma metodika kompiuterių regos ir paveikslų apdorojimo problemoms spręsti [PLQ<sup>+</sup>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, tekštūras, 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<sup>+</sup>22]), jis vis dar yra aktyviai analizuojamas, kadangi yra likusių ir pastoviai iškyylančių problemų ir iššūkių (pavyzdžiui, vaizde daug triukšmo, besikartojantys artefaktai, nelogiški piešiniai). Vaizdo žaidimų vaizdų transformavimas į realistinius vaizdus yra spręstas, 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.

Šio tyrimo tikslas ir yra sukurti modelį, kuris transformuotų nuotraukas iš realių į vaizdo žaidimų. 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. Apžvelgti modelius, tinkamus nesuporuotų vaizdų mokymui ir juos apmokyti.
3. Palyginti modelius.

# 1. Generatyviniai adversariniai tinklai

## 1.1. GAN modelis

Pirmasis generatyvinis adversarinis tinklas (angl. *generative adversarial network* – kontrastyvi neporuota transformacija) buvo pristatytas 2014 metais Montrealo universiteto mokslininkų [GPM<sup>+</sup>14]. Jo principas sudarytas iš dviejų dalių. Pirmoji dalis yra generatyvinis modelis  $G$ , kuris išmoksta nuotrauką domeno požymius ir pagal juos sugeba kurti nuotraukas. Antroji – diskriminacinis modelis  $D$ , kurio tikslas yra gavus nuotrauką nuspėti tikimybę ar nuotrauka yra iš mokomo domeno, ar yra sugeneruota modelio  $G$ . Šiuo metodu, generatyvinis modelis  $G$  yra pastoviai mokomas geriau kurti nuotraukas, kurios apgautų diskriminacinių modelių  $D$ , o diskriminacinis modelis  $D$  yra pastoviai mokomas geriau atpažinti sugeneruotą nuotrauką nuo tikros.

Šis modelis sugeba tik sukurti naujas nuotraukas, jų netransformuoja. Todėl jam reikia tik vieno duomenų rinkinio. Toliau šiame skyriuje minimi modeliai kuria nuotraukas jas transformuojant iš pirminės, todėl jiems reikalinga nuotraukos įvestis, jie taip pat reikalauja dviejų duomenų rinkinių – vieną šaltinio domeną ir vieną tikslą domeną.

## 1.2. CycleGAN modelis

Vienas iš trijų panaudotų nuotraukų transformavimo modelių yra CycleGAN modelis [ZPI<sup>+</sup>17]. Šis modelis sukurtas 2017 metais ir išpopuliarėjo dėl savo galimybės vaizdus transformuoti iš abi puses, t.y. modelių išmokius nuotraukas transformuoti iš domeno A į domeną B, jis taip pat sugebės vaizdus transformuoti iš domeno B į domeną A, nors nebūtinai taip pat gerai. Tai nulemia, jog modelių sudaro su generatyviniai modeliai ir du diskriminaciai modeliai.

Tyrimo straipsnyje yra vaizduojama, kaip modelis sugeba transformuoti vaizdus tarp realių domenų (pavyzdžiui, iš arklio į zebrą ir iš zebro į arkli) ir tarp ne nerealistinių ir realių (pavyzdžiui, iš kraštovaizdžio nuotraukų į Monet paveikslus ir atvirškčiai). Tarp dviejų nerealistinių domenų pavyzdžių nėra, tačiau galima nuspėti, jog su kokybišku apmokymo procesu ir duomenų rinkiniu tokį uždavinį taip pat nesunkiai įveiktu.

Nors modelis jau lygintinai dirbtinio intelekto pasaulyje yra senas, tačiau dėl jo kodo realizacijos prieinamumo ir architektūros paprastumo yra vertas dėmesio ir laiko.

## 1.3. CUT modelis

Antras iš trijų naudotų modelių yra CUT (angl. *contrastive unpaired translation* – kontрастyvi neporuota transformacija), išleistas 2020 metais. Pagrindinis jo bruožas yra, jog jis yra pritaikytas mokymui su nesuporuotomis nuotraukomis (t.y. nuotraukai iš domeno A, nėra tiesioginio atitikmens iš domeno B). Kitaip nei CycleGAN, CUT mokymas ir nuotraukų transformavimas vykdomas tik į vieną pusę, tai reiškia, kad norint nuotraukas transformuoti iš domeno A į domeną B ir iš domeno B į domeną A, yra reikalingi du atskirai išmokyti modeliai. Dėl vienpusio nuotraukų transformavimo, mokymo procedūra yra supaprastinama ir paspartinama. Šio modelio tikslas yra transformuojant perimti norimo domeno išvaizdą, bet išlaikyti transformuojamos nuotraukos

struktūrą ir esminį turinį. Būtent ši CUT modelio savybė ir pakiša koją transformuojant nuotraukas kai kuriuose domenuose, kadangi jei apmokant modelį yra dažnai pasitaikančią artefaktų, tai jie gali dažnai atsikartoti ir transformuojamose nuotraukose, tai yra pabrėžiama „Enhancing Photorealism Enhancement“ [RAK21] straipsnyje, kur transformuojant GTA V vaizdus, CUT modelis dažnai ant žaidėjo automobilio variklio dangčio uždėdavo Mercedes žvaigždę, kuri beveik visados matoma Cityscapes duomenų rinkinyje [COR<sup>+16</sup>], kuriuo ir buvo mokinti modeliai.

#### 1.4. MSPC modelis

Trečiasis naudotas modelis yra MSPC (angl. *Maximum Spatial Perturbation Consistency* – maksimalus erdvinės perturbacijos pastovumas) [XXW<sup>+22</sup>]. Jis yra sukurtas CycleGAN [ZPI<sup>+17</sup>] pagrindu, todėl jo nuotraukų transformavimas taip pat yra komutatyvus. Šis modelis yra sukurtas su tikslu jį naudoti nesuporuotų nuotraukų duomenų rinkiniams, kurie dažnai priveda prie nuotraukos turinio išdarkymo. Būtent šią problemą MSPC modelis ir bando spręsti, bandant geriau išsaugoti turinio bruožus ir jų turinį. Šis ir CycleGAN modeliai gali būtų mokomi suporuotais, nesuporuotais ir mišriais duomenų rinkiniais, tačiau sprendžiant šį uždavinį neįmanoma surinkti kokybiško ir kiekybiško suporuoto duomenų rinktinio, todėl yra parinktas šiuo modelių nesuporuoto mokymo metodas.

## **2. Eksperimentas**

Mokymo procesas vykdomas pagal keturis pagrindinius žingsnius:

1. Duomenų rinkinio paruošimas,
2. Modelių paruošimas;
3. Mokymas;
4. Rezultatų išvedimas.

Jei rezultatai nėra patenkinami, tai ciklą kartojama nuo numanomo problemos taško. Šiame skyriuje aprašomi šio proceso žingsniai.

### **2.1. Duomenų rinkiniai**

Vaizdų transformavimui iš realių į GTA:VC vaizdus reikia dviejų duomenų rinkinių: realaus eismo vaizdų ir žaidimo eismo vaizdų. Realaus eismo vaizdų duomenų rinkinių yra pakankamai daug, todėl jų rinkti nereikėjo ir buvo pasirinktas BDD100K[YCW<sup>+</sup>18] rinkinys, o GTA:VC vaizdams reikėjo kurti savo duomenų rinkinį, kadangi kokybiško ir kiekybiško rinkinio internete rasti nepavyko.

#### **2.1.1. Realių vaizdų duomenų rinkinys**

Kaip minėta, buvo naudojamas BDD100K duomenų rinkinys. Jis sudarytas iš 100 000 automobilių eismo nuotraukų, padarytų naudojant kameras nukreiptas į automobilio važiavimo kryptį ir yra dažniausiai ant automobilio priekinės panelės. Buvo naudotas jo poaibis sudarytas iš 10 000 nuotraukų (dar vadinamas kaip bdd10k), kadangi GTA:VC nuotraukų duomenų rinkinys susidaro lygintinai mažas. 10 tūkst. nuotraukų duomenų rinkinys yra padalintas 70:20:10 santykiu mokymui, testavimui ir validacijai (atitinkamai gaunasi 7 tūkst, 2 tūkst ir 1 tūkst), nes kitaip rinkinys gautųsi stipriai nesubalansuotas. Kiekviena nuotrauka yra 1280:720 pikselių raiškos.

#### **2.1.2. Grand Theft Auto: Vice City duomenų rinkinys**

Eksperimentas buvo vykdomas su GTA:VC vaizdais, kurių duomenų rinkinį teko sukurti. Kokybišką duomenų rinkinį leidžia sukurti žaidimo grafiniai nustatymai, todėl papildomų modifikacijų į žaidimą diegti nereikia, užtenka ekrano vaizdo įrašymo programinės įrangos. Ekrano vaizdo įrašymui buvo naudojama OBS Studio programa.

OBS Studio programos nustatyti parametrai, kad kuriamas duomenų rinkinys būtų struktūriškai kuo panašesnis į šaltinio duomenų rinkinį ir vaizdo įrašai nesigautų per dideli ir sunkiai apdrojami. Igyvendinti du parametru pakeitimai: įrašo raišką pakeista į BDD100K rinkinio nuotraukų raišką (kuri yra 1280:720 pikselių) ir nustatyta vieno kadro per sekundę įrašo sparta. Su paskutine parametru supaprastiname vaizdo įrašų perdarymą į nuotraukas, nes nereikia išmesti perteklinių

nuotraukų nustačius didesnį kadrų dažnį (pvz. nustačius standartinius 60 kadru per sekundę, gautume žymiai per daug nuotraukų).

GTA:VC žaidimas reikalauja kelių vaizdinių pakeitimų tam, kad vaizdai būtų švaresni ir struktūriškai panašesni į BDD100K duomenų rinkinį. Naudojant standartinius nustatymus yra rodomas vaizdas trečiuoju asmeniu (t.y. iš žaidėjo galio), kuriame matyti daug interfeiso elementų (žr 2 pav.). Tam, kad vaizdas būtų artimesnis šaltinio domeno duomenų rinkiniui, reikėjo panaikinti žemėlapij ir informacines detales, tą žaidimas leido nustatyti bei reikėjo nustatyti pirmą asmenį ir automobilio perspektyvos, kad nesimatytų pačio veikėjo ir taip išvengti nenorimų artefaktų (panašiai kaip išprastai lieka naudojant [COR<sup>+</sup>16] duomenų rinkinį, tame filmuojanamas vaizdas iš automobilio Mercedes, o kadangi vaizduose yra matomas automobilio kapotas prie kurio pritvirtinta ikoniška Mercedes žvaigždė, todėl daugelyje transformuotų nuotraukų atsiranda minėtoji Mercedes žvaigždė). Ši pakeitimą žaidimas taip pat leidžia daryti, kaip konfiguracinių nustatymų. Šiuos pakeitimius implementavus, gaunamas kokybiškas ir švarus vaizdas, kuris nepalieka artefaktų ir užfiksuoja esminį turinį (žr. 3 pav.).



2 pav. GTA:VC žaidimo vaizdas su išprastais nustatymais (a) ir su pakeistais nustatymais (b).

Duomenų rinkinio nuotraukos buvo kuriamos išrašinėjant GTA:VC žaidimo vaizdą, išrašuose važinėjant po žaidimo erdves keliais, bandant padengti kuo daugiau esamų vaizdų. Tas buvo daryta

tieki žaidimo dienos metu, tiek nakties, kad būtų sukuriamas duomenų rinkinio aplinkybių vienodumas ir įvairumas. Žaidimo Pasaulis yra suskirstytas į 8 regionus. Kiekvieno regiono aplinka yra nufilmuojama žaidimo dienos ir nakties metu.

Kitame skyriuje yra minima, jog rezultatuose yra pastebėta spragų – vaizdai turi žymiai mažiau automobilių nuotraukose nei BDD100K duomenų rinkinys, todėl po kelių eksperimentų, reikėjo papildyti surinktą duomenų rinkinį nuotraukomis, kuriuose yra daug automobilių arba jie užima didesnį ekrano plotą. Tokių nuotraukų iš viso buvo padaryta 300 ir jas pridėjus prie kitų nuotraukų buvo sukurta antra duomenų rinkinio versija.

Vaizdo įrašai apdoroti Python kalbos skriptu, konvertuojančiu vaizdo įrašą į atskiras kadrum nuotraukas, taip sudarydamas GTA:VC vaizdų duomenų rinkinį sudarytą iš 2000 nuotraukų, o jo antrą versiją iš 2300.

## 2.2. Mokymo platformos

Paruošus modelius, jie buvo mokomi VU MIF superkompiuteriu. Mokoma buvo naudojant Linux Ubuntu operacinę sistemą (20.04 versiją) su PyTorch karkasu, pasitelkiant superkompiuterio Nvidia DGX-1 stotį, kurią sudaro keturios Nvidia Tesla V100 vaizdo plokštės.

## 2.3. Mokymo procesas

Mokymui buvo išrinkti trys modeliai: CycleGAN<sup>1</sup>, CUT<sup>2</sup> ir MSPC<sup>3</sup>. Kadangi CUT ir MSPC kodo implementacija yra sukurta CycleGAN modelio realizacijos programinio kodo pagrindu, tiek duomenų rinkinio, tiek pačio mokymo proceso keisti drastiškai nereikėjo.

Kiekvienas modelis buvo mokomas 100 epochų su vientisu mokymo greičiu (0.0002) ir 35 epochomis tiesišku mokymosi greičio nykimu (angl. *decay*). Kiekvienas modelis buvo apmokytas du kartus, vieną kartą su pirmaja GTA:VC eismo vaizdų duomenų rinkiniu, o antrą kartą su antraja jo versija, todėl iš viso yra kiekvieno modelio dvi versijos vadinamos V1 ir V2, pavyzdžiui, CUT V1 ir CUT V2. Mokymosi greitis yra apskaičiuojama pagal šią formulę:

$$lr = 2 \times 10^{-4} \times \frac{\max(0, ep - ep_n)}{ep_d + 1},$$

čia  $ep$  – einamoji epocha,  $ep_n$  – epochų kiekis,  $ep_d$  – nykimo epochų kiekis.

CycleGAN ir MSPC modelių mokymas vyko žymiai ilgiau nei CUT modelio mokymo, dėl jų transformacijų komutatyvumo požymio, kadangi reikia nuotrauką transformuoti tiek iš domeno A į B, tiek iš B į A ir MSPC modelio mokymas truko ilgiau nei CycleGAN dėl mažesnio skaičiavimų kiekių, reikalungų minėtam vaizdo detalių palaikymui.

<sup>1</sup>CycleGAN kodo realizacija pasiekama nuoroda – <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>

<sup>2</sup>CUT modelio kodo realizacija pasiekama nuoroda – <https://github.com/taesungp/contrastive-unpaired-translation>

<sup>3</sup>MSPC modelio kodo realizacija pasiekama nuoroda – <https://github.com/batmanlab/MSPC>

## 1 lentelė. Modelių mokymo trukmės

Modelio pavadinimas	Mokymo laikas, minutės
Cut V1	1524
Cut V2	1520
MSPC V1	4136
MSPC V2	3897
CycleGAN V1	2806
CycleGAN V2	2620

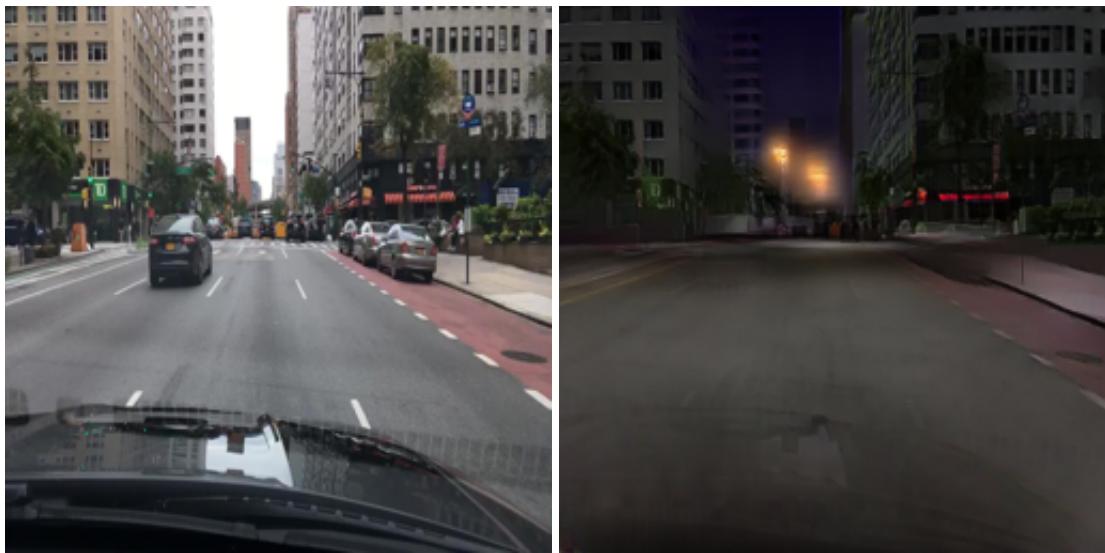
## 2.4. Modelių vertinimo metrika

Modelių nuotraukų transformavimo kokybei vertinti buvo parinkta FID (angl. *Fréchet inception distance* – Fréchet pradžios atstumas) vertinimo metrika, dažnai naudojama vertinti modelių generuojamų nuotraukų kokybei. Šia metrika apskaičiuojamas atstumas tarp realių ir sugeneruotų nuotraukų požymių vektorių.

Šis įvertis parodo, kaip nuotraukos iš dviejų grupių yra struktūriškai panašios. Žemesnis įvertis reiškia kokybiškesnes nuotraukas, o didelis – prastesnes. FID metrika pirmą kartą pristatyta ir panaudota 2017 metais Linco Johaneso Keplerio universiteto tyrimo straipsnyje pavadinimu "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium"[HRU<sup>+</sup>17]. Šis įvertis pristatytas kaip geresnė metrika vietoj pradžios įverčio (angl. *inception score*, trumpinama IS), kadangi IS vertina tik sugeneruotų nuotraukų kokybę, tačiau FID vertina sugeneruotų nuotraukų kokybę lyginant su šaltinio domeno duomenų rinkiniu.

## 2.5. „Dingstančių automobilių“ problema

Išmokius CycleGAN modelį buvo pastebėta, jog pradėjo dingti tam tikros stambios detalės iš nuotraukų, svarbiausia iš jų – automobiliai. Išmokyta pirmoji CycleGAN modelio versija linksta trinti automobilius ir uždengti tai aplinkos detalėmis ar fonu. Dėl to nuspresta buvo išbandyti CUT modelį, kadangi jis pritaikytas nesuporuotų duomenų rinkinių uždaviniams, tačiau pirmoji jo versija yra dar labiau linkusi trinti automobilius iš nuotraukų. Dėl šios problemos taip pat buvo išbandytas MSPC modelis, nes juo yra sprendžiama nuotraukų semantinių detalių palaikymo problema. Nors MSPC modelis pasiekia geresnius rezultatus nei kiti du ties šia problema, tačiau išlaikomų automobilių kiekis vis tiek liko labai mažas (žr. 2 lentelę).



3 pav. CycleGAN modelio transformuotos nuotraukos pavyzdys, kuriame yra panaikinami automobiliai.

2 lentelė. Atpažintų automobilių kiekis duomenų rinkiniuose.

Modelio pavadinimas	Atpažintų automobilių kiekis, vnt.	Atpažintų automobilių dalis, %
Testavimo duomenų rinkinys	426	-
CycleGAN V1	15	3,5 %
CycleGAN V2	14	3,2 %
Cut V1	7	1,6 %
Cut V2	10	2,3 %
MSPC V1	10	2,3 %
MSPC V2	17	3,9 %

Automobilių atpažinties nuotraukose testavimui naudojamas Centernet HG-104 [DBX<sup>+22</sup>] modelis iš anksto apmokytas su Microsoft CoCo duomenų rinkiniu [LMB<sup>+14</sup>]. Šis modelis sugeba pakankamai gerai atpažinti automobilius net ir vaizdo žaidimuose kaip GTA:VC.

Modeliai išmokytį su patobulintu duomenų rinkiniu, kuriame yra daugiau nuotraukų, kuriose yra automobilių užimančių pakankamai didelį nuotraukos plotą, arba yra nuotraukų su daug automobilių nuotraukoje, demonstruoja geresnį sugebėjimą išlaikyti automobilius (išskyrus su CycleGAN modeliu, kurio statistika minimaliai sumažėjo). Atpažintų automobilių padidėjimas siekia 69 % su MSPC modeliu, tačiau tokis patobulėjimas yra beveik bereikšmis kadangi bendras atpažintų automobilių procentas lieka labai mažas.

Gali būti daug priežasčių, kodėl modeliai yra taip linkę trinti automobilius iš nuotraukų, tačiau labiausiai tikėtina iš jų yra automobilių kiekiai duomenų rinkiniuose – GTA:VC duomenų rinkinyje jų yra žymiai mažiau. 3 lentelėje yra pateikti mokymo duomenų rinkiniuose esantys automobilių kiekiai ir kokia dalis nuotraukų turi juose atpažintą bent vieną automobilį.

3 lentelė. Duomenų rinkinių atpažintų automobilių kiekiai ir nuotraukų.

Duomenų rinkinys	Atpažintų automobilių kiekis, vnt.	Dalis nuotraukų kuriuose yra atpažinta automobilių, %
BDD100K	21624	89 %
GTA:VC V1	274	11 %
GTA:VC V2	653	23 %

## 2.6. Paros meto keitimosi problema

Gautose nuotraukose pastebėta problema, jog tam tikrais atvejais yra pakeičiamas paros metas iš dienos į naktį. Tas dažniausiai įvyksta kai būna debesuota arba kai yra mažai matoma dangaus, tarkim kai aplinkui yra daug aukštų pastatų arba gamtos (žr. 5 pav.).



4 pav. CycleGAN V2 pavyzdžiai kur modelio dėl aukštų pastatų arba debesų yra diena pakeičiamama į naktį.

## 2.7. Gauti FID įverčiai

Apmokius modelius buvo apskaičiuotos FID metrikos indikuojančios, kaip kokybiskai modeliai sugeba transformuoti nuotraukas lyginant su tikslu domenu.

4 lentelė. Apmokyti modelių gauti FID įverčiai ir atpažintų automobilių kiekiai.

Modelio pavadinimas	FID	Atpažintų automobilių kiekis, vnt.
CycleGAN V1	170	15
CycleGAN V2	167	14
Cut V1	181	7
Cut V2	186	10
MSPC V1	167	10
MSPC V2	145	17

Remiantis 4 lentelės duomenimis, matoma, jog geriausias FID nebūtinai reiškia geriausius rezultatus. Antroji apmokyta CycleGAN versija patobulino savo FID, tačiau, nors ir minimaliai, nukrito atpažistamų automobilių skaičius, o CUT modeliui yra atvirkštinė situacija – antroji modelio versija padidino aptiktų automobilių skaičių, tačiau dėl to sumažėjo FID įvertis. Vienintelis modelis visiškai patobulėjęs nuo pagerinto GTA:VC duomenų rinkinio yra MSPC, jo atpažintų automobilių skaičius padidėjo ir FID įvertis sumažėjo. Pagal šiuos rezultatus matoma, jog iš trijų bandytų modelių, MSPC teikia geriausius rezultatus.

## Kokybiškai transformuoojamos detalės

Modeliai itin gerai sugeba transformuoti gamtinius objektus, kaip žolė ir medžiai, iš realaus domeno į vaizdo žaidimo domeną. GTA:VC duomenų rinkinyje yra daug žaidime paplitusių palmių, kurios atrodo maždaug vienodai ir yra pavaizduotos 6 paveiksluke.



5 pav. GTA:VC domeno palmių pavyzdys.

Modelis medžius sugeba gerai atskirti ir juos transformuoti į GTA:VC domene matomus. Iš tamsiai žalios į šviesiai žalią spalvą pereinantys lapai yra nupiešiami medžiams ir net kamienai yra transformuojami į dryžuotai rudus. Ši elgsena yra būdinga visiems trims modeliams ir abejoms kiekvieno jų versijoms.



6 pav. MSPC V2 pavyzdžiai kur modelio dėl aukštų pastatų arba debesų yra diena pakeičiama į naktį.

Kitos dvi gerai transformuojamos detalės yra kelias ir dangus (kelio ir dangaus pavyzdys matomas 6 pav.). Dangus beveik visada išgauna išprastą saulėtos dienos mėlyną arba giedros nakties juodą spalvą. Nors, kaip prieš tai minėta, dangus kartais pakeičia dienos laiką, tačiau ir naktinis dangus yra tiksliai transformuojamas. Asfalto spalva taip pat yra labai gerai parenkama. Nors šios detalės yra labai paprastai transformuojamos, tačiau jos sudaro didelę dalį bendros transformuotos nuotraukos kokybės ir panašumo į vaizdo žaidimo domeną.

# **Rezultatai ir išvados**

Gausesni transformuotų nuotraukų pavyzdžiai yra prisegti priedų skiltyje

## **FID įverčiai ir automobilių atpažintų automobilių kiekiai**

Modelių sąrašas nuo geriausio iki blogiausio vertinant pagal FID:

1. MSPC V2
2. CycleGAN V2
3. MSPC V1
4. CycleGAN V1
5. CUT V1
6. CUT V2

Modelių sąrašas nuo geriausio iki blogiausio vertinant pagal transformuotose nuotraukose atpažintų automobilių kiekius:

1. MSPC V2
2. CycleGAN V1
3. CycleGAN V2
4. MSPC V1
5. CUT V2
6. CUT V1

## **Išvados**

Vaizdų transformavimas buvo įgyvendintas iki tokio lygio, kad galima nesunkiai atpažinti transformuoto vaizdo domeną, tačiau yra vietų, kurias tobulinti reikėtų, pagrindinė iš jų: nuotraukų semantinio turinio išlaikymas (paros metas, automobiliai). Modeliai dažnai panaikina automobilius iš transformuojamų nuotraukų, tačiau gerai sugeba transformuoti kelius, dangų ir gamtinius objektus. Tyrimo išvados:

1. Kokybėskiausius rezultatus duodantis modelis yra MSPC.
2. Duomenų rinkinių struktūriniai ir objektų dažnumo skirtumai stipriai įtakoja nuotraukų transformavimo kokybę.
3. Kokybėskesnis nuotraukos transformavimas nebūtinai reiškia kokybėskesnį nuotraukos turinio išlaikymą.

## **Santrumpos**

1. **GTA:VC** – Žaidimas „Grand Theft Auto: Vice City“.
2. **GTA V** – Žaidimas „Grand Theft Auto 5“.
3. **FID** – angl. *Fréchet inception distance* – Fréchet pradžios atstumas, yra metrika naudojama įvertinti generatyvinių modelių kokybę.
4. **IS** – angl. *Inception score* – pradžios įvertis, yra metrika naudojama įvertinti generatyvinių modelių kokybę.

## Literatūra

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

## **Priedas nr. 1**

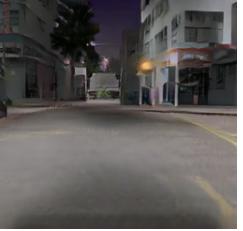
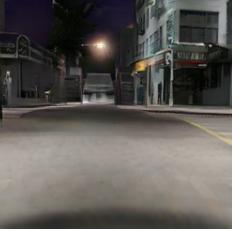
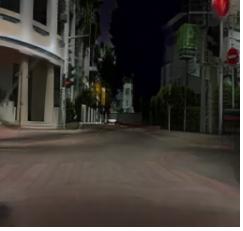
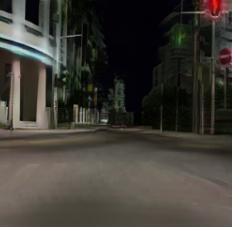
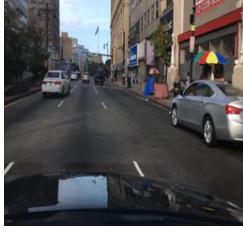
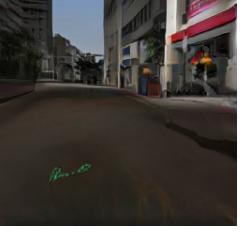
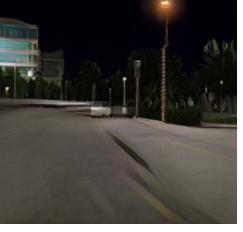
### **Darbo saugykla**

Darbo saugykla su duomenų rinkiniu, pavyzdinėmis nuotraukomis ir gautais modeliais yra pasiekiamą nuorodą:  
<https://github.com/0ksas/kursinis>.

## Priedas nr. 2

### Nuotraukų pavyzdžiai (pirmosios duomenų rinkinio iteracijos)

5 lentelė. Transformuojamų nuotraukų ir jų rezultatų pavyzdžiai

Originali nuotrauka	CycleGAN	CUT	MSPC
			
			
			
			
			
			

### Priedas nr. 3

#### Nuotraukų pavyzdžiai (antrosios duomenų rinkinio iteracijos)

6 lentelė. Transformuojamų nuotraukų ir jų rezultatų pavyzdžiai

Originali nuotrauka	CycleGAN V2	CUT V2	MSPC V2
			
			
		