Kauno technologijos universitetas

Informatikos fakultetas

Generatyvinės duomenų augmentacijos tyrimas dviejų plokštumų monokuliniam nelambertinių paviršių gylio žemėlapių sudarymui

Magistro baigiamasis projektas

Vilius Bankauskas

Projekto autorius

Prof. dr. Andrius Kriščiūnas

Vadovas

Kaunas, 2026

Kauno technologijos universitetas

Informatikos fakultetas

Generatyvinės duomenų augmentacijos tyrimas dviejų plokštumų monokuliniam nelambertinių paviršių gylio žemėlapių sudarymui

Magistro baigiamasis projektas

Dirbtinio intelekto informatika (6211BX007)

Vilius Bankauskas
Projekto autorius

Prof. dr. Andrius Kriščiūnas
Vadovas

Dr. Dalia Čalnerytė
Recenzentė

Kaunas, 2026

Kauno technologijos universitetas

Informatikos fakultetas

Vilius Bankauskas

Generatyvinės duomenų augmentacijos tyrimas dviejų plokštumų monokuliniam nelambertinių paviršių gylio žemėlapių sudarymui

Akademinio sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;

2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąraše;

3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);

4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Vilius Bankauskas

Patvirtinta elektroniniu būdu

Bankauskas Vilius. Generatyvinės duomenų augmentacijos tyrimas dviejų plokštumų monokuliniam nelambertinių paviršių gylio žemėlapių sudarymui. Magistro baigiamasis projektas / projektui vadovavo Prof. dr. Andrius Kriščiūnas; Kauno technologijos universitetas, informatikos fakultetas.

Studijų kryptis ir studijų krypčių grupė: Informatikos mokslai, Informatika (B01).

Reikšminiai žodžiai: monokuliarinis gylio nustatymas, nelambertiniai paviršiai, duomenų augmentacija.

Kaunas, 2026. 72 p.

Santrauka

Baigiamajame projekte nagrinėjama viena aktualiausių ir sparčiausiai evoliucionuojančių šiuolaikinės kompiuterinės regos sričių – monokuliarinis gylio nustatymas (angl. Monocular Depth Estimation, MDE), kuris yra kritiškai svarbus autonominių sistemų, robotikos bei papildytos realybės technologijų vystymui. Pagrindinė mokslinė ir inžinerinė problema, kurią siekiama spręsti darbe, kyla dėl fundamentalių fizikinių ribojimų: dabartiniai populiarūs gylio nustatymo algoritmai bei fiziniai jutikliai, tokie kaip LiDAR ar „Time-of-Flight“ kameros, dažnai nesugeba korektiškai identifikuoti skaidrių, pusiau permatomų ar šviesą atspindinčių paviršių geometrijos. Kadangi šie nelambertiniai objektai neatspindi šviesos tolygiai, jie iškraipo jutiklių matavimus, todėl standartiniai neuroniniai tinklai klaidingai interpretuoja jų atstumą nuo kameros arba laiko juos tuščia erdve. Projekto tyrimo objektas apima monokuliarinio gylio nustatymo modelių architektūras, jų mokymo procesus bei gautų gylio žemėlapių tikslumą analizuojant būtent nelambertinius paviršius. Darbo tikslas – pagerinti gylio nustatymo modelių gebėjimą patikimai vertinti skaidrius objektus, pasitelkiant inovatyvų metodą – generatyvinio dirbtinio intelekto sukuriamus aukštos kokybės sintetinius mokymo duomenis. Siekiant tikslo, darbe buvo įgyvendinti keli esminiai uždaviniai: atlikta išsami mokslinės literatūros analizė apie modernias MDE architektūras, sukurta unikali duomenų sintezės metodika pasitelkiant multimodalinį modelį „Google Gemini“ (kuris leido simuliuoti skaidrių objektų vaizdus kartu su jų geometrinėmis savybėmis fone), bei atliktas pasirinkto bazinio „DepthAnythingV2“ modelio adaptavimas ir apmokymas (angl. fine-tuning). Tyrimo metu taikyti sisteminės literatūros apžvalgos, vaizdų apdorojimo, neuroninių tinklų mokymo bei lyginamosios analizės metodai. Vertinant modelių patobulinimus, buvo naudojamas „DIODE“ duomenų rinkinys bei kiekybinės metrikos, tokios kaip RMSE, AbsRel ir tikslumo koeficientas δ (delta). Projekto realizacijos metu gauti rezultatai atskleidė, kad integruoti sintetiniai duomenys leidžia MDE algoritmams žymiai geriau atskirti skaidrius objektus nuo fono geometrijos, o tai tiesiogiai koreliuoja su sumažėjusiomis prognozavimo klaidomis ir padidėjusiu gylio žemėlapių vizualiniu vientisumu sudėtingose scenose. Galiausiai, išvadose konstatuojama, kad generatyvinis dirbtinis intelektas yra efektyvus įrankis trūkstamų duomenų spragoms užpildyti, o tai atveria naujas galimybes tikslesniam aplinkos suvokimui. Darbo struktūrą sudaro įvadas, santrumpų bei terminų sąrašas, literatūros apžvalga, metodologinė dalis, realizacijos ir eksperimentų skyrius, išvados bei literatūros šaltinių sąrašas.

Bankauskas Vilius. Generative Data Augmentation "Research" for Dual-Plane Monocular Depth Estimation of Non-Lambertian Surfaces. Master‘s Final Degree Project / supervisor Prof. Dr. Andrius Kriščiūnas; Faculty of Informatics, Kaunas University of Technology.

Study field and study field group: Computer science, Informatics (B01).

Keywords: monocular depth estimation, non-lambertian surfaces, data augmentation.

Kaunas, 2026. 72 pages.

Summary

This project examines one of the most relevant and rapidly evolving fields of modern computer vision – Monocular Depth Estimation (MDE), which is critically important for the development of autonomous systems, robotics, and augmented reality technologies. The primary scientific and engineering problem addressed in this work stems from fundamental physical limitations: currently popular depth estimation algorithms and physical sensors, such as LiDAR or Time-of-Flight cameras, often fail to correctly identify the geometry of transparent, semi-transparent, or reflective surfaces. Since these non-Lambertian objects do not reflect light uniformly, they distort sensor measurements, causing standard neural networks to misinterpret their distance from the camera or treat them as empty space. The research object of this project encompasses monocular depth estimation model architectures, their training processes, and the accuracy of the resulting depth maps when analyzing specifically non-Lambertian surfaces. The aim of the work is to improve the ability of depth estimation models to reliably evaluate transparent objects by employing an innovative method – high-quality synthetic training data created by generative artificial intelligence. To achieve this goal, several key objectives were implemented: a comprehensive scientific literature analysis of modern MDE architectures was performed, a unique data synthesis methodology was developed using the multimodal model Google Gemini (which allowed for the simulation of transparent object images along with their geometric properties in the background), and the adaptation and fine-tuning of the selected base DepthAnythingV2 model were executed. During the research, methods of systematic literature review, image processing, neural network training, and comparative analysis were applied. To evaluate the model improvements, the DIODE dataset and quantitative metrics such as RMSE, AbsRel, and the accuracy coefficient delta were used. The results obtained during the implementation of the project revealed that integrated synthetic data allow MDE algorithms to significantly better distinguish transparent objects from the background geometry, which directly correlates with reduced prediction errors and increased visual consistency of depth maps in complex scenes. Finally, the conclusions state that generative artificial intelligence is an effective tool for filling missing data gaps, which opens new possibilities for more accurate environmental perception. The structure of the work consists of an introduction, a list of abbreviations and terms, a literature review, a methodological part, an implementation and experimentation section, conclusions, and a list of references.

Turinys

Lentelių sąrašas ... 8
Paveikslų sąrašas ... 9
Santrumpų ir terminų sąrašas ... 11
Įvadas ... 12
1. Monokuliarinio gylio nustatymo ir nelambertinių paviršių literatūros analizė ... 14
1.1. Duomenų generavimas ir augmentacija ... 15
1.1.1. Sintetinių aplinkų naudojimas ir domeno poslinkis ... 15
1.1.2. Generatyvinių ir difuzinių modelių taikymas duomenų augmentacijai ... 16
1.1.3. Alternatyvių jutiklių duomenų simuliacija ... 17
1.2. Sprendimai nelambertiniams paviršiams ... 17
1.2.1. Architektūrinės modifikacijos ir kelių logikų modeliai ... 17
1.2.2. Fizika ir geometrija paremti metodai ... 18
1.2.3. Skirtingų jutiklių apjungimas (angl. Sensor Fusion) ... 19
2. Generatyvinės duomenų augmentacijos metodika ... 20
2.1. Projekto apimtis ... 20
2.1.1. Sintetiniai duomenys 1 (ignoruojami skaidrūs objektai) ... 20
2.1.2. Sintetiniai duomenys 2 (matomi skaidrūs objektai) ... 20
2.2. Duomenų paruošimas ... 21
2.3. Mokymo hiperparametrai ... 22
2.4. Vertinimo kriterijai ... 22
2.4.1. Tikslumo metrikos ... 23
2.4.2. Klaidos metrikos ... 23
3. Duomenų sintezė ... 25
3.1. Generatyvinių modelių apribojimai ir kokybės kriterijai ... 25
3.2. Sintetinių duomenų analizė ... 31
3.2.1. Scenų variacija ... 36
3.3. Duomenų analizės rezultatai ... 38
4. Monokuliarinio gylio modelių eksperimentiniai rezultatai ... 40
4.1. Eksperimentų aplinka ir sąlygos: techninė įranga programinė įranga ... 40
4.1.1. Techninė įranga ... 40
4.1.2. Programinė įranga ... 40
4.2. Duomenų rinkinio statistika ... 41
4.2.1. Pirmojo plano duomenų rinkinys (angl. Foreground dataset) ... 41
4.2.2. Fono duomenų rinkinys (angl. Background dataset) ... 42
4.3. Mokymas: eiga, mokymosi kreivės ir rezultatų aptarimas ... 42
4.3.1. Mokymo dinamika ... 42
4.3.2. Duomenų kiekio įtaka ... 42
4.3.3. Mokymosi kreivės: fono gylio žemėlapių modeliai ... 43
4.3.4. Mokymosi kreivės: pirmo plano gylio žemėlapių modeliai ... 48
4.4. Rezultatų analizė ... 52
4.4.1. Gylio žemėlapių kokybės įvertinimas standartinėse scenose ... 56
4.4.2. Specializuotų modelių įvertinimas originaliame domene ... 59
4.4.3. Specializuotų modelių įvertinimas realaus pasaulio aplinkoje ... 63
4.5. Modelių pritaikymas ... 65

Išvados 67
Dirbtinio intelekto įrankių naudojimas 68
Literatūros sąrašas 69

Lentelių sąrašas

1 lentelė. Mokymo hiperparametrai ir jų pasirinkimo priežastys 22
2 lentelė. Asmeninio kompiuterio sudedamosios dalys 40
3 lentelė. Programinė įranga 41
4 lentelė. Fono aptikimo modelių duomenų pasiskirstymas 43
5 lentelė. Pirmo plano aptikimo modelių duomenų pasiskirstymas 43
6 lentelė. Klaidos metrikų įverčiai 57
7 lentelė. Tikslumo metrikų įverčiai 58
8 lentelė. Dešimt modifikuotų nematytų scenų metrikų įverčiai 60

Paveikslų sąrašas

1 pav. TODD (Toronto Transparent Objects Depth Dataset) rinkinio pavyzdžiai 15
2 pav. Sugeneruotos nuotraukos naudojant skirtingas sudėtingų oro sąlygų instrukcijas 16
3 pav. Scenos rekonstrukcijos generavimo srautas 17
4 pav. Trisluoksnio gylio nustatymo modelio dizaino išklotinė 18
5 pav. Kokybinis gylio žemėlapių palyginimas naudojant albedo žemėlapius 19
6 pav. Siūlomos augmentacijos proceso blokinė schema, originalus duomenų rinkinys paverčiamas į du lygiagrečius augmentuotus duomenų rinkinius 20
7 pav. Objekto sąveikos su šviesa skirtumas transformuojant medžiagą į permatomą. Pabrėžiamos naujai atsiradusios refrakcijos, vidiniai atspindžiai ir kaustikos (iliustracija sugeneruota naudojantis Gemini 3 Pro Image modelį) 25
8 pav. SDInpainting ir ResNetFPN modelių klaidingai sugeneruotų segmentacijos kaukių ir modifikuotų scenų pavyzdžiai 27
9 pav. ResNetFPN modelio nesėkmingos segmentacijos kaukės pavyzdys 27
10 pav. SDInpainting ir ResNetFPN modelių segmentacijos kaukės ir modifikuojamos scenos 28
11 pav. SDXL ir ResNetFPNV2 modelių segmentacijos kaukės ir modifikuojamos scenos 29
12 pav. FLUX modelio skaidraus objekto pridėjimo į sceną pavyzdžiai 30
13 pav. FLUX modelio medžiagos transformavimo pavyzdžiai 31
14 pav. Scenų, kuriuose generavimo procesas atliktas sėkmingai, pavyzdžiai 32
15 pav. Scenos pavyzdžiai, kuriuose pridedant naują skaidrų kūną kartu pridedamas ir neskaidrus objektas 33
16 pav. Medžiagos transformacijos pavyzdžiai, kuriuose bendra scenos geometrija išlaikyta, tačiau sukuriama transformuoto kūno geometrijos klaida 33
17 pav. Klaidingai konvertuoto kūno scenos pavyzdys. Sienos dekoro objekto transformacija 34
18 pav. Medžiagos transformavimo užduoties įvykdymas scenai neturint aptinkamų objektų 35
19 pav. Medžiagos transformavimo klaida susiduriant su mažais objektais 36
20 pav. Skaidraus objekto įklijavimo scenos variacijos pavyzdys 37
21 pav. Medžiagos transformavimo scenos variacijos pavyzdys 37
22 pav. Fono aptikimui sugeneruotų duomenų įvertinimų pasiskirstymo skritulinė diagrama 38
23 pav. Pirmo plano aptikimui sugeneruotų duomenų įvertinimų pasiskirstymo skritulinė diagrama 39
24 pav. Matytų duomenų strategijos fono aptikimo modelių mokymosi kreivė 45
25 pav. Nematytų duomenų strategijos fono aptikimo modelių mokymosi kreivės 46
26 pav. Hibridinės strategijos fono aptikimo modelių mokymosi kreivės 47
27 pav. Matytų duomenų strategijos pirmo plano aptikimo modelių mokymosi kreivė 49
28 pav. Nematytų duomenų strategijos pirmo plano aptikimo modelių mokymosi kreivės 50
29 pav. Hibridinės strategijos pirmo plano aptikimo modelių mokymosi kreivės 51
30 pav. Visų aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija 53
31 pav. Fono aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija 54
32 pav. Pirmo plano aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija 55
33 pav. Modifikuotų scenų pavyzdžiai, kuriuose labiausiai pastebimas sudarytų gylio žemėlapių skirtumas 61
34 pav. Modifikuotų scenų pavyzdžiai, kuriuose visi modeliai prognozavo fono, o ne stiklo plokštumos, gylio žemėlapius 62

35 pav. DIODE duomenų rinkinio scena su dideliu kiekiu nelambertinių paviršių. Skirtingų modelių gylio žemėlapiai radikaliai skiriasi vienas nuo kito ir nuo tikrojo jutiklio gauto atstumo 63
36 pav. Realaus pasaulio scena su taure ir stalu (Nuotaukos šaltinis: [49]) 63
37 pav. Realaus pasaulio scenos su stiklo plokšte ir stikline (Nuotaukų šaltiniai: [50, 51]) 64
38 pav. Realaus pasaulio scenarijus su dviem stikliniais buteliais (Nuotraukos šaltinis: [52])..... 64
39 pav. Realaus pasaulio scenarijus su rudomis ir skaidriomis ampulėmis (Nuotraukos šaltinis: [53]) 64
40 pav. Realaus pasaulio scena su skaidriu buteliu ir delnu (Nuotraukos šaltinis: [54])..... 65
41 pav. Palyginus sugeneruotus gylio žemėlapius naudojantis pirmo ir antro plano modelius, gaunama binarinė segmentacijos kaukė padengianti skaidrius butelius (Nuotraukos šaltinis: [52]) 66

Santrumpų ir terminų sąrašas

AbsRel – absoliutus santykinis skirtumas (angl. Absolute Relative Difference).

AR – papildyta realybė (angl. Augmented Reality).

DI – dirbtinis intelektas (angl. Artificial Intelligence).

LiDAR – lazerinis vaizdo aptikimo ir atstumo nustatymo metodas (angl. Light Detection and Ranging).

Log10 – vidutinė logaritminė dešimtainė paklaida (angl. Mean log10 error).

MDE – monokuliarinis gylio nustatymas (angl. Monocular Depth Estimation).

Nelambertiniai paviršiai – paviršiai, kurie neatspindi šviesos tolygiai visomis kryptimis (pvz., veidrodiniai, stikliniai).

RGB – spalvų modelis, sudarytas iš raudonos, žalios ir mėlynos spalvų (angl. Red, Green, Blue).

RMSE – vidutinė kvadratinė paklaida (angl. Root Mean Square Error).

SqRel – kvadratinis santykinis skirtumas (angl. Squared Relative Difference).

δ (delta) – tikslumo metrika, rodanti prognozuotų taškų dalį, kurios santykinė paklaida neviršija nustatytos ribos.

Įvadas

Projekto naujumas ir aktualumas

Monokuliarinis gylio nustatymas (angl. Monocular depth estimation, MDE) yra viena iš sparčiausiai besivystančių kompiuterinės regos sričių, kurioje, pasitelkiant neuroninius tinklus, iš vienos 2D nuotraukos yra sukuriamas scenos 3D gylio žemėlapis. Nors šiuolaikiniai dirbtinio intelekto modeliai demonstruoja aukštus rezultatus standartinėse situacijose, jie susiduria su kritiniais iššūkiais analizuojant nelambertinius – skaidrius (stiklas, skaidrus plastikas) ar atspindinčius (veidrodžiai, poliruoti paviršiai) – kūnus. Tokie objektai neturi savitų vizualinių bruožų, o jų išvaizda priklauso nuo aplinkos atspindžių ar fono, todėl modeliai dažnai juos interpretuoja klaidingai, pavyzdžiui, kaip tuščią erdvę.

Šio darbo aktualumas pasireiškia siekiu spręsti šią fundamentalią kompiuterinės regos problemą, kuri riboja autonominių sistemų, robotikos bei papildytos realybės technologijų patikimumą realaus pasaulio sąlygomis. Kadangi tikslių gylio duomenų surinkimas skaidriems objektams naudojant fizinius jutiklius yra itin sudėtingas, brangus ir dažnai netikslus procesas, kyla poreikis ieškoti alternatyvių mokymo duomenų šaltinių. Projekto naujumas slypi specializuoto sintetinių duomenų rinkinio, orientuoto išskirtinai į pirmojo plano skaidrius objektus ir už jų esančio fono geometriją, sukūrime, panaudojant generatyvinio dirbtinio intelekto technologijas. Šis požiūris leidžia išvengti tradiciniams kompiuterinės grafikos (CGI) metodams būdingų trūkumų ir atveria naujas galimybes adaptuoti bazinius gylio nustatymo modelius sudėtingoms scenoms.

Tikslas ir uždaviniai

Tyrimo objektas: Monokuliarinio gylio nustatymo (MDE) modelių pritaikymas atpažįstant arba ignoruojant skaidrius ir atspindinčius (nelambertinius) paviršius.

Tyrimo tikslas: Pagerinti monokuliarinio gylio nustatymo modelių veikimą dirbant su skaidriais objektais, pasitelkiant generatyvinio dirbtinio intelekto sugeneruotus sintetinius mokymo duomenis.

Tyrimo uždaviniai:

1. Išanalizuoti mokslinę literatūrą, susijusią su duomenų generavimo ir monokuliarinio gylio nustatymo sprendimais nelambertiniams paviršiams.
2. Sukurti specializuotą sintetinių duomenų rinkinį pasitelkiant generatyvinius modelius, simuliuojančius pirmojo plano skaidrius objektus ir už jų esančio fono geometriją.
3. Apmokyti bazinius gylio nustatymo modelius atpažinti arba ignoruoti skaidrius objektus, naudojant originalų „DIODE“ ir naujai sugeneruotą sintetinį duomenų rinkinius.
4. Atlikti apmokytų modelių analizę, vertinant gylio žemėlapių tikslumą ir daromas paklaidas standartinėse bei realaus pasaulio scenose.

Tyrimo metodai ir modeliai: Mokslinės literatūros analizė; sintetinių duomenų generavimas pasitelkiant dirbtinio intelekto modelius; giluminių neuroninių tinklų apmokymas; lyginamoji kiekybinė analizė naudojant klaidų (RMSE, AbsRel, SqRel, Log10) bei tikslumo (δ) metrikas; kokybinė vizualinė gautų gylio žemėlapių rezultatų analizė.

Dokumento struktūra

Darbą sudaro įvadas, keturi pagrindiniai skyriai, išvados ir literatūros sąrašas. Pirmame skyriuje atliekama mokslinės literatūros analizė, kurioje apžvelgiami duomenų generavimo, augmentacijos metodai ir specifiniai architektūriniai sprendimai dirbant su nelambertiniais paviršiais. Antrame skyriuje pristatoma tyrimo metodologija – detalizuojama sintetinių duomenų paruošimo logika dviem skirtingoms plokštumoms (fonui ir pirmam planui), aprašomi modelių mokymo parametrai bei naudojamos vertinimo metrikos. Trečiame skyriuje aprašomas praktinis duomenų sintezės procesas naudojant difuzinius bei multimodalinius modelius, aptariami generavimo apribojimai ir atliekama gautų duomenų kokybės analizė. Ketvirtame skyriuje pateikiami apmokytų gylio nustatymo modelių eksperimentiniai rezultatai, analizuojama mokymo dinamika bei atliekamas detalus modelių vertinimas standartinėse ir realaus pasaulio scenose.

1. Monokuliarinio gylio nustatymo ir nelambertinių paviršių literatūros analizė

Skyriuje apžvelgiama mokslinė literatūra ir naujausi tyrimai, susiję su monokuliariniu gylio nustatymu. Siekiant nuosekliai pagrįsti vėlesniuose darbo skyriuose priimamus praktinius sprendimus, analizė struktūrizuojama atspindint pagrindinius projekto etapus ir yra dalijama į dvi esmines dalis. Pirmoje dalyje analizuojami duomenų generavimo bei augmentacijos metodai, kuriais siekiama susidoroti su kokybiškų mokymo duomenų trūkumu. Antroje dalyje dėmesys skiriamas specifiniams modelių architektūriniams sprendimams bei logikoms, taikomoms dirbant su sudėtingais, nelambertiniais kūnais.

Monokuliarinis gylio nustatymas (angl. Monocular depth estimation, MDE) yra sparčiai besivystanti kompiuterinės regos sritis, kurioje pritaikoma ir dirbtinio intelekto technologija. MDE yra kompiuterinės regos uždavinys, kuriame, pasitelkiant neuroninius tinklus, yra sudaromas scenos 3D gylio žemėlapis, remiantis viena 2D RGB nuotrauka. Tai aktualu daugelyje skirtingų pritaikymo sričių nuo robotų navigacijos iki papildytos realybės (angl. augmented reality). Pastaraisiais metais padidėjo mokslinio pasaulio susidomėjimas šia sritimi dėl neuroninių tinklų architektūrų patobulėjimo ir palengvėjusio kokybiškų duomenų pasiekiamumo. Aukšto tikslumo jutikliai, tokie kaip LiDAR, kurie geba išgauti reikalingus gylio žemėlapius iš scenų, būdavo skirti tik specializuotam personalui dėl savo kainos ir prieinamumo. Sumažėjus jutiklių kainoms ir LiDAR jutiklių integravimas į populiarų išmanųjį telefoną Apple iPhone 12 Pro, palengvino gylio duomenų rinkimo procesą [1]. Duomenų prieinamumas suteikė sąlygas spartesniam modelių tobulėjimui [2].

MDE uždaviniui spręsti yra skiriami daugiau nei keli atviros prieigos modeliai, pasižymintys aukštais kokybiniais įvertinimais [3]. Depth Anything, UniDepth ir MiDaS yra vieni iš populiariausių modelių, kurių dėka MDE sritis plėtėsi ir progresavo [4, 5, 6]. Pastarieji geba suformuoti tikslius gylio žemėlapius sudėtingose ir objektų pripildytose scenose. Nors modeliai sugeba susitvarkyti su natūraliai dažnai pasikartojančiomis scenomis ir objektais, tačiau susiduria su problemomis uždavinio ir realaus pasaulio scenų struktūrose. Monokuliarinis gylio nustatymas yra korektiškai nesuformuluotas uždavinys (angl. ill-posed), kadangi viena 2D scenos projekcija gali turėti begalybę 3D scenų atitikmenų [7].

Kita problema kyla iš natūraliai susiformuojančių scenų sudėties. Kasdieniai objektai, tokie kaip durys, sienos ir grindys, yra dažnai pastebimi duomenų rinkiniuose ir gylio žemėlapiuose, todėl modeliai išmoksta jų vizualias savybes ir patikimai sudaro jų gylio žemėlapius. Tačiau ne visi kasdieniai objektai yra tokie paprasti. Kai kurios durys turi stiklo komponentų, kitos, pavyzdžiui, spintos durys, gali turėti įmontuotą veidrodį. Scenose dažni objektai, kaip stiklinės, langai, plastikiniai buteliai ar dieninės užuolaidos egzistuoja beveik kiekviename kambaryje, o jų išvaizda nevisiškai atspindi jų tikrąją formą ar atstumą nuo kameros. Objektui esant didžiąja dalimi permatomam, yra susiduriama su objekto vaizdinių užuominų trūkumu, tokiu atveju toks objektas, kaip stiklinės durys, yra vizualiai apgaulingas ir beveik neatskiriamas nuo tuščios erdvės. Žmogui, pamačius minimalius šviesos atspindžius ir ore kabančią durų rankeną, iš karto aišku, kad tai yra stiklinės durys. Objektai, kurie atspindi šviesą, kelia dar didesnį iššūkį; gerai prižiūrėti, neturintys ryškių bruožų (angl. featureless) paviršiai dažnai geba tobulai atspindėti sceną, kuri fiziškai neegzistuoja, o tai stipriai iškraipo jutiklių gaunamus duomenis.

Bandymai susidoroti su nelambertiniais kūnais dažnai remiasi gylio užpildymu (angl. depth completion) technikomis [8, 9] ir sintetinėmis CGI aplinkomis [10]. Sintetinės aplinkos gerai

matematiškai simuliuoja skaidrių objektų sąveiką su šviesos spinduliais, tačiau šie metodai dažnai susiduria su sintetinių aplinkų specifine išvaizda. Tai sukuria duomenis, kurie neatrodo fotorealistiškai [11]. Modeliai mokyti šiais sintetiniais duomenimis išmoksta atpažinti generavimo artefaktus labiau negu tikrąsias stiklo, skaidraus plastiko ar veidrodžio vizualines užuominas [12, 13, 14]. Tai sukelia domeno poslinkį modeliams susiduriant su realaus pasaulio scenomis.

Pastaraisiais metais matomas aiškus mokslinis poslinkis link sintetinių duomenų augmentacijos siekiant padidinti sunkiai gaunamų duomenų kiekį [15-18]. Sintetinė duomenų augmentacija taip pat turi apribojimų - sugeneruotų duomenų kokybė tiesiogiai priklauso nuo generuojančio metodo kokybės.

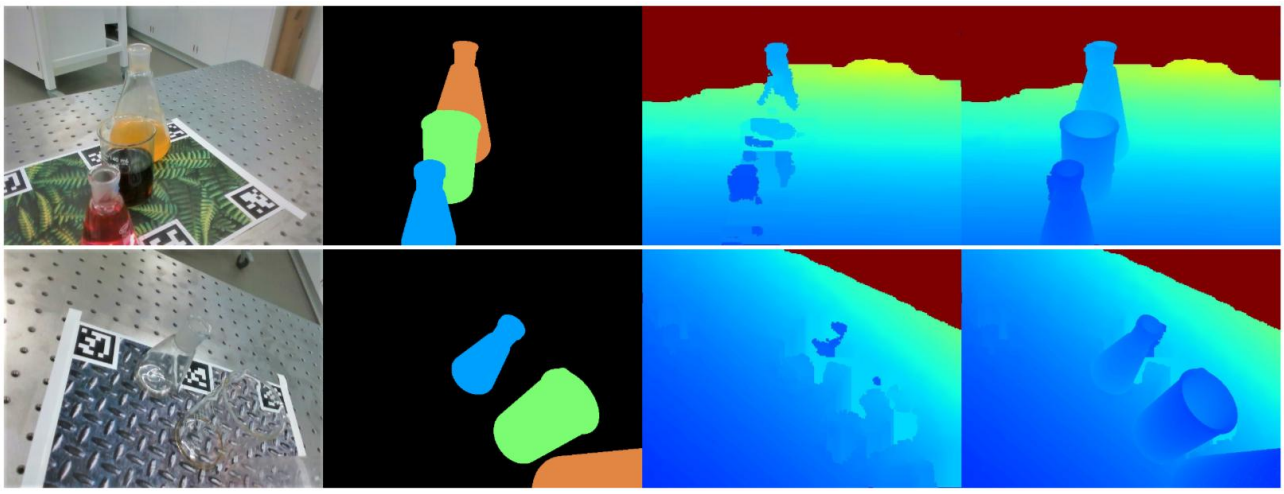
1.1. Duomenų generavimas ir augmentacija

Tikslių gylio duomenų, ypač atspindintiems ar skaidriems paviršiams, surinkimas yra itin brangus, sudėtingas ir rankinio darbo reikalaujantis procesas. Todėl mokslinėje literatūroje didelis dėmesys skiriamas alternatyviems duomenų gavimo metodams.

1.1.1. Sintetinių aplinkų naudojimas ir domeno poslinkis

Siekdami išspręsti mokymo duomenų stygiaus problemą, tyrėjai dažnai pasitelkia sintetinių duomenų generavimą. Vienas tokių būdų yra generuoti fotorealistiškus vaizdus kompiuterinių žaidimų grafikos variklių aplinkose, taip išvengiant realybėje pasitaikančių defektų [19]. Sukūrus vaizdų imtį su įvairiomis oro sąlygomis ir apšvietimu, išsprendžiama kiekio problema, tačiau susiduriama su domeno poslinkio (angl. domain shift) iššūkiu – modeliai, apmokyti su sintetiniais duomenimis, sunkiai generalizuoja realaus pasaulio vaizdus

Domeno poslinkio problema ypač išryškėja dirbant su skaidriais objektais. Egzistuojantys skaidrių kūnų duomenų rinkiniai yra per „paprasti“ - neturi sudėtingų fonų ar kokybiškos gylio informacijos. Siekdami tai išspręsti, tyrėjai pateikia TODD (Toronto Transparent Objects Depth Dataset) duomenų rinkinį (žr. **1 pav.**), kuriame užfiksuota 15 000 RGB-D vaizdų realiose sudėtingose scenose [8]. Visgi, autorių taikomas taškų debesies užpildymo (PCC) metodas generuoja per retus rezultatus, o dėl skaičiavimo apkrovų tankių žemėlapių sugeneruoti nepavyksta, ypač kai scenoje vienas skaidrus objektas uždengia kitą.



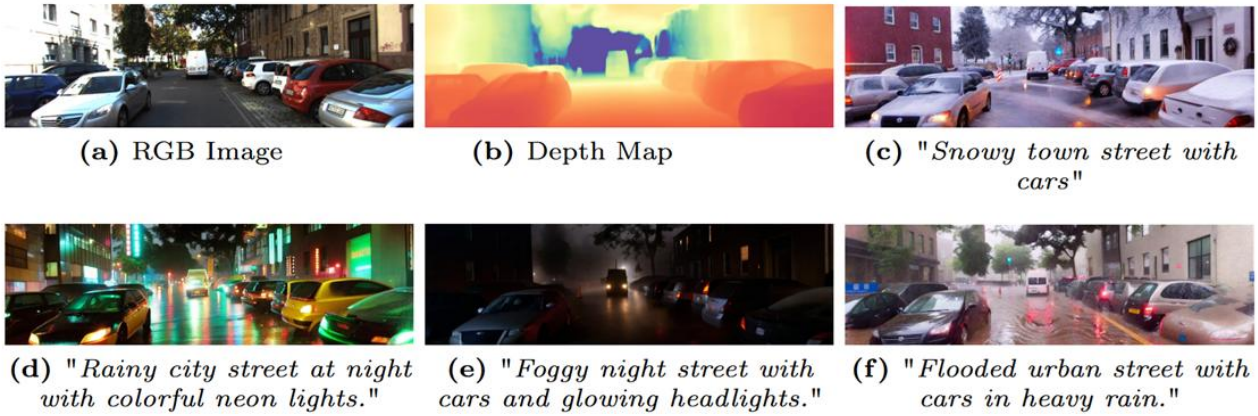
1 pav. TODD (Toronto Transparent Objects Depth Dataset) rinkinio pavyzdžiai

Duomenų stokos ir nesudėtingumo problemai spręsti yra taikomi skirtingi generavimo varikliai gebantys generuoti didelius duomenų rinkinius. Tyrėjai, pasitelkę naudojant fizika paremtą „Blender Cycles“ variklį, geba sukurti daugiau nei 50 000 RGB-D vaizdų sintetinį rinkinį [9]. Testavimo patikimumui užtikrinti, realaus pasaulio duomenys renkami rankiniu metodu – skaidrūs objektai po nufotografavimo keičiami identiškomis, bet dažais padengtomis kopijomis, kad jutiklis galėtų nuskaityti tikslų paviršiaus gylį. Tokia kombinacija padidino modelio atsparumą.

Vis dėlto, dirbtinis paviršių modifikavimas turi ribų. „Domeno atsitiktinumo“ (angl. domain randomization) technika yra svarbi trinant ribą tarp simuliacijos ir realybės [10]. Šio metodo sėkmė priklauso nuo milžiniškų duomenų kiekių, o esant per dideliam medžiagų savybių atsitiktinumui, modelis praranda gebėjimą generalizuoti. Tai dar kartą patvirtina tyrimas, kurio metu „Blender Cycles“ pagalba sugeneruoti dvigubi gylio žemėlapiai (fonui ir skaidriam paviršiui) vis tiek lėmė žymiai prastesnius modelio rezultatus pereinant į realaus pasaulio scenas [11].

1.1.2. Generatyvinių ir difuzinių modelių taikymas duomenų augmentacijai

Generatyviniai dirbtinio intelekto, ypač difuziniai, modeliai literatūroje vis dažniau taikomi siekiant įveikti nestandartinių (angl. out-of-distribution, OOD) duomenų trūkumo problemas [20]. Dirbant su sudėtingomis sąlygomis – naktinėmis scenomis ar stiklu – tradicinis duomenų surinkimas yra dažnai neįmanomas. Vienas iš siūlomų sprendimų: paprastos scenos gylio žemėlapis perduodamas į „ControlNet“ tinklą kaip ribojimas, ir, naudojant tekstines užklausas („skaidrus stiklinis objektas“, žr. **2 pav.**), sugeneruojamas naujas vaizdas su sudėtingomis tekstūromis, bet išsaugota pradine fizine geometrija [21]. Tai leidžia praplėsti duomenų rinkinį nerenkant duomenų rankiniu metodu.



2 pav. Sugeneruotos nuotraukos naudojant skirtingas sudėtingų oro sąlygų instrukcijas

Kitas pažangus požiūris, kuriame tekstiniai difuzijos modeliai apjungiami su vizualinės kalbos modeliais (VLM), siekiant sugeneruoti natūraliai atrodančius, bet optiškai klaidinančius 3D objektus [22]. Nors iš pradžių tai kurta kaip ataka prieš autonominio vairavimo sistemas, autoriai pabrėžia tokio generavimo vertę kuriant augmentuotas mokymo bazes, kurios gali būti naudojamos duomenų rinkinio praplėtimui ir modelio atsparumo susiduriant su klaidinančiais paviršiais gerinimui.

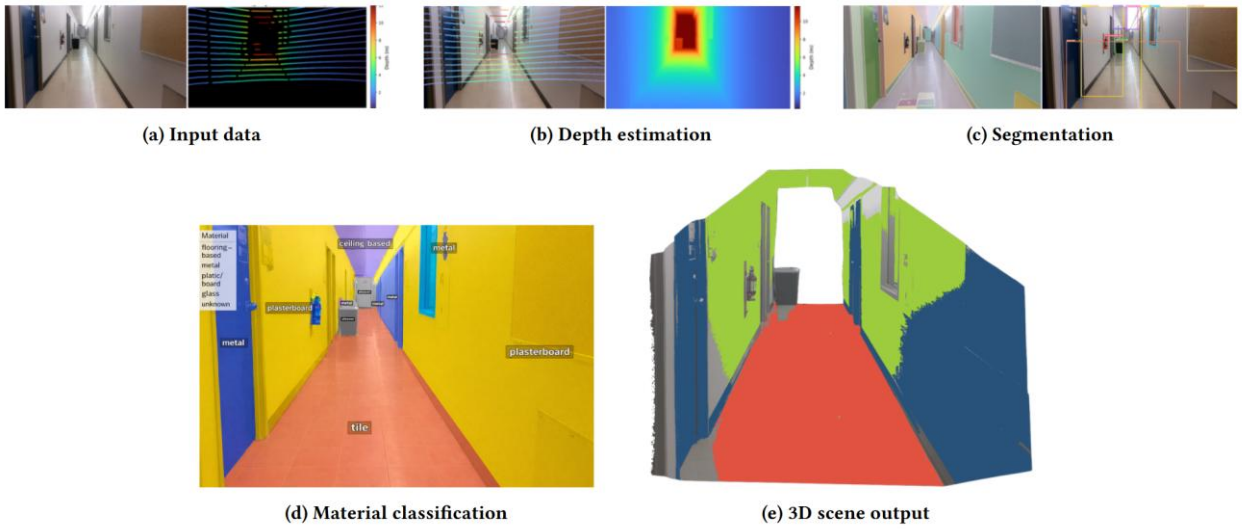
Nepaisant pastarųjų metodų inovacijos, jie turi esminių trūkumų. Pirmiausia, tai klaidų paveldimumas: kadangi baziniai gylio žemėlapiai kuriami modelių, jų daromos klaidos pereina į naują mokymo rinkinį. Antra, difuzijos procesas kartais sukuria „haliucinacijas“ (smulkius objektus, neatitinkančius gylio žemėlapio). Galiausiai, net ir pažangiausi generatyviniai objektai, perkeliami į

realias gatvės scenas, praranda savo efektyvumą dėl klasikinio simuliacijos-realybės domeno poslinkio.

1.1.3. Alternatyvių jutiklių duomenų simuliacija

Susiduriant su RGB ir LiDAR jutiklių trūkumais, atsigręžiama į radaro technologijas [23]. Radaro jutiklių gaunami duomenys pasižymi aukšta gylio kokybe ir atsparumu blogam matomumui, tačiau dėl mažos matavimo zonos bei ilgo nuskaitymo laiko jiems trūksta didelės apimties duomenų rinkinių. Inovatyvus metodas sprendžia šią problemą generuojant milimetrinių bangų (mmWave) radaro mokymo duomenis tiesiogiai iš 2D nuotraukų (žr. **3 pav.**) [24].

Sistema suskirsto vaizdą į objektus ir naudoja vizualinės kalbos modelius (VLM) jų fizinėms medžiagoms prognozuoti, dėl to automatiškai sukuriami konfigūruojami radaro atspindžiai. Tačiau autoriai pastebi, kad taip sugeneruoti taškų debesys sudaro tik apie 12 % realaus radaro duomenų tankio. Be to, kadangi simuliacija priklauso nuo siauro kameros matymo kampo, ji nesugeneruoja atspindžių objektams, esantiems už kadro ribų, o paties metodo efektyvumas kol kas įrodytas tik su viena objektų klase (durimis).



3 pav. Scenos rekonstrukcijos generavimo srautas

1.2. Sprendimai nelambertiniams paviršiams

Antroje literatūros analizės dalyje dėmesys skiriamas specifiniams modelių architektūriniams sprendimams, vertinant atspindinčius ir skaidrius objektus, kai duomenų augmentacijos nepakanka.

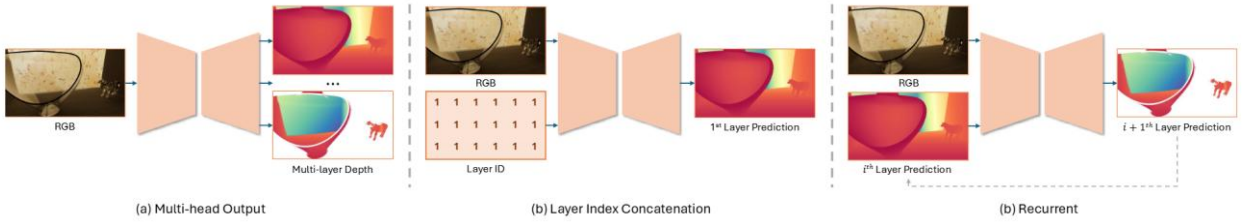
1.2.1. Architektūrinės modifikacijos ir kelių logikų modeliai

Susidūrę su skaidriais kūnais, standartiniai monokuliarinio gylio modeliai patiria dvi logines problemas: tinklas privalo pasirinkti, ar vertinti atstumą iki paties skaidraus objekto paviršiaus, ar iki fono, esančio už jo. Tokia dviprasmybė sukelia drastiškas gylio vertinimo klaidas.

Siekdami tai išspręsti, tyrėjai, taiko architektūrinį sprendimą su sintetiniais duomenimis: skaidrūs kūnai segmentuojami, nustatomos paviršiaus normalės ir ribos [25]. Integruojant šiuos parametrus,

modelis identifikuoja „vizualiai apgaulingas“ zonas ir joms pritaiko maksimalių bei minimalių gylio verčių ribojimus.

Kitas siūlomas sprendimas pristatoma „Multi-Layer Depth Anything (MLDA)“ architektūra (žr. **4 pav.**) [11]. „DepthAnythingV2“ modelio pagrindu sukurtas tinklas turi dvi galvas: viena aptinka skaidrų paviršių, o kita jį ignoruoja ir įvertina fono gylį. Nors dviejų verčių vienam pikseliui prognozavimas yra sudėtingas ir jautrus domeno poslinkiui, tai esminis žingsnis link kompleksinio scenos suvokimo.



4 pav. Trisluoksnio gylio nustatymo modelio dizaino išklotinė

Kiti tyrimai, bando modifikuoti bazinius modelius, įvedant naujas klaidų funkcijas ir vaizdų apdorojimo modulius [6]. Nors tai pagerina skaidrių kūnų aptikimą, modelis vis tiek turi dideles paklaidas stipraus apšvietimo sukeltam tekstūrų praradimui. Panaši problema identifikuojama sprendžiant gylio dviprasmybę be tiesioginio modelio treniravimo (angl. zero-shot). Siūlomas naujas Laplaso operatoriumi (LVP) paremtas metodas, išryškinantis aukšto dažnio detales gylio sluoksniams atskirti [26]. Deja, metodas žlunga, jei scenoje trūksta kontrasto arba paviršius itin stipriai atspindi šviesą.

Pažymėtina, kad į architektūrą vis dažniau integruojami ir difuziniai mechanizmai [27]. Tačiau tokie sprendimai dažnai susiduria su persimokymo problema (angl. overfitting) remiantis paviršiaus tekstūromis ir reikalauja neproporcingai didelių skaičiavimo resursų.

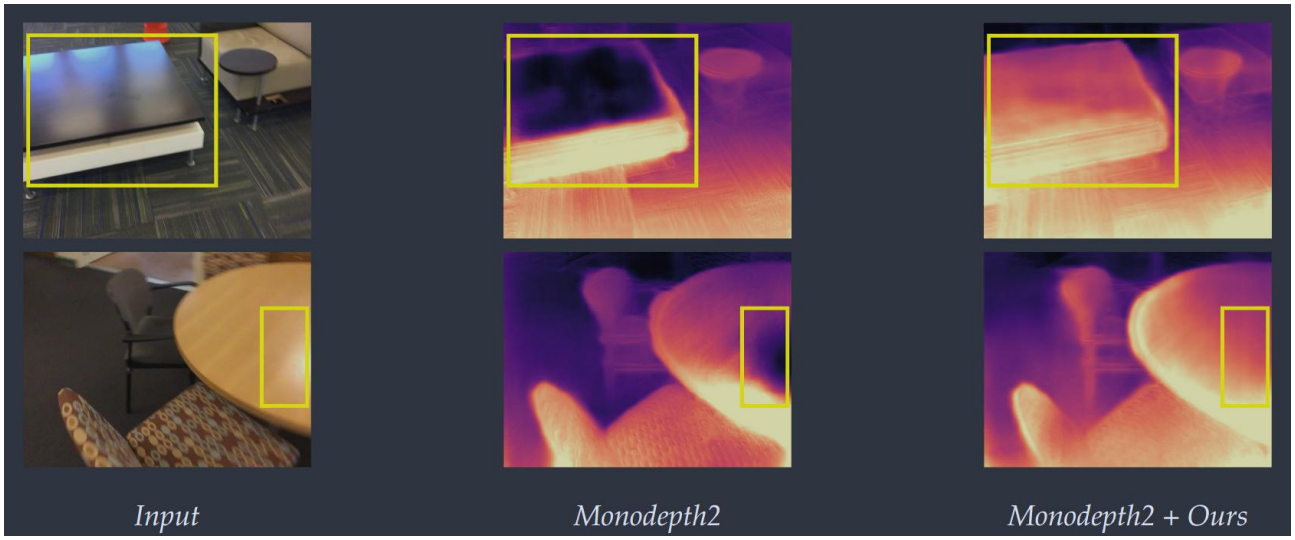
Skirtingiems metodams palyginti organizuojami konkursai, kuriuose modeliai lyginami pagal savo gebėjimą susidoroti su skaidriais paviršiais [28]. Tačiau net galingiausi baziniai modeliai, apmokyti su dešimtimis milijonų vaizdų, vis dar daro kritines klaidas vertindami nelambertinius paviršius (ypač vandenį).

1.2.2. Fizika ir geometrija paremti metodai

Dalis mokslininkų problemą sprendžia remdamiesi fizikos dėsniais ir erdvės geometrija, dažnai taikydami neprižiūrimąjį mokymą (angl. unsupervised learning). Naudojamos nuotraukų sekos ir remiamasi foto metrinės dermės (angl. photometric consistency) prielaida [29]. Nors tai leidžia atsisakyti ground truth duomenų, prielaida žlunga susidūrus su nelambertiniais kūnais, nes keičiantis apžvalgos kampui keičiasi jų spalva ir atspindžiai. Bandymai tai kompensuoti modelio kaukėmis stipriai priklauso nuo pačio segmentacijos modelio tikslumo.

Siekiant atsparumo apšvietimui ir atspindžiams, siūlomas inovatyvus sprendimas, pagrįstas objektų albedo (vidine, nuo apšvietimo nepriklausančia spalva) [30]. Naudojant vidinio vaizdo dekompozicijos tinklą (IID), pradinė nuotrauka padalijama į albedo ir šešėlių sluoksnius. Kadangi atspindžiai pažeidžia lambertinę prielaidą, modelis remiasi stabiliais albedo duomenimis (žr. **5 pav.**).

Nors šis metodas puikiai veikia su šlapiais paviršiais ar metalu, stiklo atveju IID tinklas vis tiek nesugeba atskirti stiklo paviršiaus nuo už jo esančio vaizdo.



5 pav. Kokybinis gylio žemėlapių palyginimas naudojant albedo žemėlapius

Kita alternatyva yra geometriniai erdvės atskaitos taškai. Sistema „Murre“ naudoja retą iš struktūros judesyje (SfM) gautą taškų debesį kaip inkarą (angl. guide) monokuliariniam gylio tinklui [31]. Tai atleidžia modelį nuo spėliojimų dėl mastelio. Visgi, metodas paveldi visus SfM trūkumus: susidūrus su atspindžiais, skaidriais ar judančiais objektais, SfM algoritmas žlunga, kartu nutraukdamas ir tolesnę scenos gylio rekonstrukciją.

1.2.3. Skirtingų jutiklių apjungimas (angl. Sensor Fusion)

Suvokiant fundamentalius vien vizualinės informacijos ribotumus dirbant su skaidriais kūnais, literatūroje tiriamas skirtingų fizinių jutiklių apjungimas. Milimetrinėms bangoms (mmWave) skaidrios medžiagos atrodo kaip nepermatomos, todėl radarai yra natūralus sprendimas kylančiai problemai.

„FuseGrasp“ sistema sujungia radaro ir kameros (RGB-D) duomenis, siekiant padėti robotinėms rankoms tiksliau manipuliuoti skaidriais objektais [32]. Roboto ranka juda virš objektų ir nuosekliai skenuoja erdvę, kurdama radaro vaizdus, kurie gilaus mokymosi tinkle apjungiami su kameros informacija. Sistema išlaiko stabilumą net prasto apšvietimo sąlygomis ir sėkmingai aptinka stiklą.

Tačiau metodas atskleidžia naujų praktinių barjerų: radaro vaizdo kokybė tiesiogiai priklauso nuo fizinio roboto rankos judesio tikslumo ir lėto skenavimo greičio. Sumažėjus rankos tikslumui, radaro vaizdai susilieja. Nors kameros ir radaro sintezė yra labai perspektyvi, jos pilną potencialą šiuo metu stipriai riboja lėtas radaro mokymo duomenų surinkimo procesas.

2. Generatyvinės duomenų augmentacijos metodika

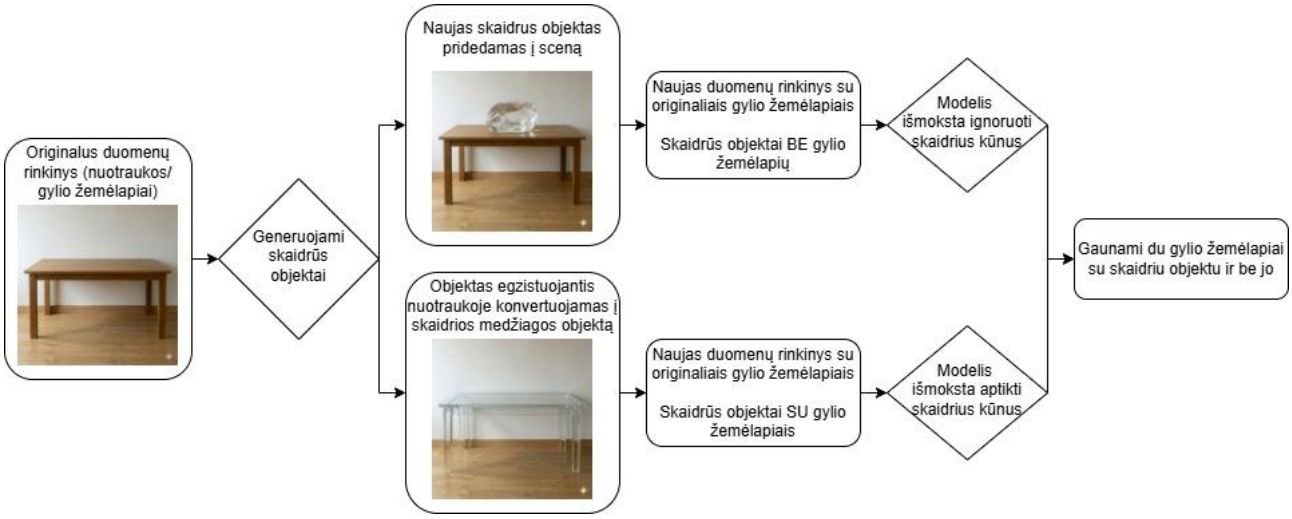
2.1. Projekto apimtis

Pagrindinis duomenų sintezės tikslas yra augmentuoti jau egzistuojantį RGB-D duomenų rinkinį taip, kad duomenų rinkinys būtų konvertuojamas į du skirtingo tipo duomenų rinkinius (6 pav.):

1. aptinkami skaidrūs objektai (pirmas planas);

2. ignoruojami skaidrūs objektai (antras planas).

Tai pasiekiama naudojant generatyvinį dirbtinį intelektą: „įdažant“ (angl. inpaint) sceną, papildant ją naujais skaidriais objektais, kurie originaliai neegzistavo scenoje ir konvertuojant jau egzistuojančius objektus į skaidrius kūnus, siekiant kuo labiau išlaikyti originalų objekto geometrinį identitetą. Atlikus augmentaciją, sukuriami du skirtingi duomenų rinkiniai: pirmas rinkinys su naujais skaidriais kūnais, kurie neegzistuoja gylio žemėlapyje; antras - su objektais, kurie egzistuoja gylio žemėlapyje, tačiau dabar yra vizualiai skaidrūs, remiantis nuotraukos duomenimis.



6 pav. Siūlomos augmentacijos proceso blokinė schema, originalus duomenų rinkinys paverčiamas į du lygiagrečius augmentuotus duomenų rinkinius

2.1.1. Sintetiniai duomenys 1 (ignoruojami skaidrūs objektai)

Panaudojus difuzinius modelius generuoti naujus objektus scenoje nekeičiant gylio žemėlapio duomenų, mokomas modelis priverstas susidurti su neatitikimais tarp vaizdo ir gylio duomenų. Dėl šių naujų objektų permatomumo, informacija, kuri būtų prarasta (uždengta naujo objekto), tampa stipriai pakeista, tačiau suteikia galimybę aproksimuoti už objekto egzistuojančią sceną. Logikos veikimo principas panašus į žmogaus regą, kur svarbi informacija apie už objekto esančią sceną nėra ignoruojama, o panaudojama formuojant projekciją [33, 34].

2.1.2. Sintetiniai duomenys 2 (matomi skaidrūs objektai)

Transformuojant jau egzistuojantį objektą į permatomą arba pusiau permatomą medžiagą sukuriamas labiau standartinis ir įprastas duomenų rinkinys, kuriame skaidrūs objektai yra matomi kaip bet kokie kiti objektai ir niekuo neišskiriami. Tokie naujai konvertuoti objektai yra visiškai matomi gylio žemėlapiuose, todėl išlaikius jų formą, jie visiškai nesiskiria nuo aplinkos. Tai suteikia unikalų būdą

gauti skaidrių objektų gylio žemėlapius neatliekant naujų matavimų jutikliais ir išvengiant nelambertinių kūnų iškeliamų problemų. Viso to pasekoje, modelį galima apmokyti iš daug aukštos kokybės ir tikslumo sunkiai natūraliai gaunamų duomenų su tiksliais gylio žemėlapiais ir pernaudojant jau egzistuojančius duomenų rinkinius.

Sintezės logikos pagrindas – siekis išspręsti fundamentalią regos ir gylio jutiklių problemą, atsirandančią dėl nelambertinių paviršių savybių. Standartiniai jutikliai dažniausiai nepajėgia tiksliai užfiksuoti skaidrių medžiagų gylio, o natūralių tokių objektų duomenų surinkimas ir rankinis anotavimas tampa itin sudėtingu bei brangiu procesu [35, 36]. Todėl pagrindinis generavimo srautų tikslas yra sukurti lengvai naudotiną ir kontroliuojamą mokymo aplinką, kuri leistų peržengti fizinių jutiklių matavimų ribas.

Suteikiant modeliams prieigą prie dviejų kontrastingų scenarijų, modelis yra priverstas išmokti skirtingas skaidrumo interpretacijas: vienu atveju – gebėti filtruoti skaidrius kūnus kaip optinius trikdžius ir atkurti fono geometriją, kitu – tiksliai lokalizuoti patį skaidrų objektą tridimensinėje erdvėje. Toks požiūris į duomenų augmentaciją ne tik išspręstų anotuotų duomenų trūkumo problemą, bet ir padėtų išugdyti giliojo mokymosi modelių gebėjimą išnaudoti daugiau gaunamos vizualinės informacijos.

2.2. Duomenų paruošimas

Tyrimui atlikti pasirinktas DIODE duomenų rinkinys [37]. Pagrindinė DIODE duomenų rinkinio pasirinkimo priežastis yra didelė scenų nuotraukų raiška (1024x768). Didelės raiškos nuotraukų naudojimas suteikia daugiau duomenų augmentacijai ir atspindi realistiškesnius scenarijus. Tai aktualu difuziniams modeliams, kurie modifikuoja vaizdinius duomenis. Didesnės rezoliucijos nuotraukos išvengia situacijų, kur transformuojamas objektas turi mažai vaizdinių duomenų ir užtikrina aukštesnę sugeneruotų objektų kokybę.

NYU Depth V2 ir kiti duomenų rinkiniai yra aktyviai naudojami sprendžiant MDE uždavinį [38]. DIODE duomenų rinkinys buvo pasirinktas dėl savo plataus kambarių scenų (angl. „indoors“) pasirinkimo ir aukštos kokybės lazeriu skenuotų gylio duomenų. Priešingai nuo kitų duomenų rinkinių, kurie remiasi žemesnės raiškos gylio žemėlapiais, DIODE rinkinys suteikia tankius gylio žemėlapius uždarose erdvėse ir pasiekia netgi 99,6 % jutiklio spindulių grįžimo lygį. Aukštas spinduliuotės grįžimas suteikia duomenų rinkiniui patikimumo ir leidžia modeliams išvengti neteisingų arba klaidinančių duomenų mokymosi etape.

Sugeneruotos nuotraukos kokybė taip pat priklauso nuo originalios nuotraukos. Kadangi kai kurios scenos DIODE duomenų rinkinyje yra be objektų, kuriuos būtų galima konvertuoti į skaidrius, susiduriama su konvertavimo problema. Papildomai aktualu paminėti, kad DIODE duomenų rinkinys buvo sudaromas atliekant pilną patalpos nuskaitymą su FARO Focus S350 lazeriniu jutikliu ant trikojo stovo. Jutiklis atlieka vieną sferinį skenavimą ir apima 360 laipsnių horizontaliai ir 150 vertikaliai. Gauti duomenys buvo išskaidyti DIODE tyrimo komandos į nuotraukų (1024x768) ir gylio žemėlapių duomenų rinkinį. Tačiau atliekant tokio tipo patalpos skenavimą, gautame duomenų rinkinyje dažnai susiduriama su „tuščiomis“ scenomis. Scenose aptikti tokie objektai kaip sienos ir grindys ar lubos, tačiau jokių kitų objektų nėra matoma. Tai sukelia nestandartines sąlygas generatyviniam modeliui, kuris priverstas „išlaikyti scenos struktūrą“, tačiau pakeisti scenoje egzistuojantį objektą. Tokios situacijos nėra viena kitai trukdančios, tačiau susiduriant su scenomis, kuriose nėra objektų, kurie galėtų būti konvertuojami, modelis yra priverstas arba atlikti radikalius

pokyčius scenoje, arba nieko nepakeisti. Siekiant to išvengti, duomenų rinkinį galima papildomai filtruoti pagal nuotraukos vertikalų laipsnį, siekiant užtikrinti kuo didesnį į kadrą patenkančių objektų kiekį, tačiau tai apribotu duomenų rinkinį ir paveiktų modelio mokymo ir veikimo rezultatus.

Duomenų rinkinys DIODE yra metrinio gylio duomenų rinkinys. DepthAnythingV2 yra santykinio gylio modelis. Metrinio gylio duomenų rinkinio pavertimas į santykinį yra dažnai taikoma praktika [3, 39, 40]. Pats populiariausias būdas transformuoti metrinį gylį į santykinį yra atvirkštinio gylio metodas (angl. inverse depth / disparity). Gylio žemėlapio vertės invertuojamos ir tolimesni objektai artėja link nulinės reikšmės, o artimi objektai turi didžiausią reikšmę iki maksimalios reikšmės 1.

2.3. Mokymo hiperparametrai

Skyriuje aprašomi DepthAnythingV2_S modelio treniravimui naudoti hiperparametrai. Pasirinktų reikšmių priežastys dokumentuotos **1 lentelė**.

1 lentelė. Mokymo hiperparametrai ir jų pasirinkimo priežastys

Parametras	Vertė/reikšmė	Priežastis
split_ratio	0.8	Standartinis padalinimas užtikrina didžiausią duomenų kiekį mokymui, kartu išlaikant pakankamą imtį objektyviai validacijai.
target_size	(1024, 768)	Vaizdai keičiami į standartizuotą raišką siekiant išlaikyti aukštą detalumą, kuris būtinas tiksliam gylio įvertinimui. (Originali DIODE nuotraukų raiška)
batch_size	4	Itin mažas partijos (angl. batch) dydis pasirinktas siekiant išvengti vaizdo plokštės atminties trūkumo (OOM) dirbant su dideliais modeliais ir aukštos raiškos vaizdais. Su pasirinkta verte pasiektas greičiausias treniravimo greitis.
Lr	5e-6	Mažas mokymosi greitis parinktas tam, kad būtų atsargiai tobulinami iš anksto ištreniruoto (angl. pre-trained) modelio svoriai jų nesugadinant.
max_epochs	50	Epochų apribojimas leidžia mokymo metu pritaikyti ankstyvojo stabdymo funkcija, kad procesas būtų nutrauktas pasiekus tenkinamą vertę.
Patience	10	Ankstyvojo stabdymo riba nutraukia treniravimą po dešimties epochų be validacijos klaidos sumažėjimo, taip efektyviai apsaugant nuo persimokymo (angl. overfitting).
Optimizer	AdamW	Optimizatorius pasirinktas dėl efektyvaus svorių slopinimo (angl. weight decay) mechanizmo, kuris yra standartinis ir patikimas pasirinkimas PyTorch transformerinių modelių treniravimui.
loss_function	scale_invariant_loss	Pritaikyta L1 funkcija normalizuoja prognozes ir tikruosius duomenis, ir suteikia galimybę modelio mokymo metu naudoti santykinių gylio skirtumų, o ne absoliučių atstumų.

Papildomai buvo naudojamas tik vienas duomenų krovėjo darbuotojas (angl. worker). Vertė pasirinkta išspręsti duomenų krovėjo metamas klaidas veikiant Microsoft Windows 11 programinės įrangos aplinkoje.

2.4. Vertinimo kriterijai

Siekiant teisingai interpretuoti darbo rezultatus pasirinktos vertinimo metrikos. Visas gylio nustatymo metrikas galima padalinti į dvi pagrindines skiltis: tikslumo metrikos ir klaidos metrikos.

Visose formulėse naudojami žymėjimai:

- $\boldsymbol{d_p}$ – modelio prognozuotas gylis
- $\boldsymbol{d_t}$ – tikras gylis
- $\boldsymbol{N}$ – visų vertinamų pikselių (taškų) skaičius

2.4.1. Tikslumo metrikos

Metrika $\boldsymbol{\delta < 1.25}$ yra viena iš pagrindinių tikslumo metrikų naudojamų vertinant gylio nustatymo modelius [6, 41, 42]. Ji nurodo, kokia dalis modelio spėjimų buvo labai arti tikrosios vertės. Kadangi tai yra tikslumo metrika, didesnė jos reikšmė reiškia tikslesnius modelio rezultatus.

Kiekviename nuotraukos taške (pikselio koordinatėse) modelis prognozuoja atstumą ($\boldsymbol{d_p}$). Iš duomenų rinkinio žinomas tikrasis to taško atstumas ($\boldsymbol{d_t}$). Metrika lygina šias dvi reikšmes ieškodama jų santykio (žr. 1 formulė). Siekiant teisingai įvertinti tiek pervertintą, tiek nepakankamai įvertintą atstumą, visada skaičiuojamas didžiausias galimas šių dviejų skaičių santykis:

$$Slenkstis = max\left(\frac{d_p}{d_t}, \frac{d_t}{d_p}\right) \quad (1)$$

Metrika $\boldsymbol{\delta < 1.25}$ skaičiuoja procentinę dalį visų pikselių, kuriuose santykis yra mažesnis nei 1.25:

$$\delta_1 = \%\ pikselių, kur\ max\left(\frac{d_p}{d_t}, \frac{d_t}{d_p}\right) < 1.25 \quad (2)$$

Slenkstis < 1.25 reiškia, kad modelio padaryta santykinė klaida yra ne didesnė nei 25 %.

Egzistuoja $\boldsymbol{\delta_2, \delta_3}$ kurios veikia lygiai tuo pačiu principu, bet leidžia didesnę paklaidą.

$\boldsymbol{\delta_2}$: slenkstis yra $1.25^2 \sim 1.56$. Modelio prognozės užskaitomos kaip teisingos, jei klaida ne daugiau kaip ~56 %:

$$\delta_2 = max\left(\frac{d_p}{d_t}, \frac{d_t}{d_p}\right) < 1.25^2 \quad (3)$$

$\boldsymbol{\delta_3}$: slenkstis yra $1.25^3 \sim 1.95$. Modelio prognozės užskaitomos kaip teisingos, jei klaida ne daugiau kaip ~95 %:

$$\delta_3 = max\left(\frac{d_p}{d_t}, \frac{d_t}{d_p}\right) < 1.25^3 \quad (4)$$

Paprastai analizuojant modelius $\boldsymbol{\delta_1}$ yra svarbiausia, nes ji geriausiai atspindi griežtą modelio tikslumą,

2.4.2. Klaidos metrikos

AbsRel (Absoliuti santykinė klaida):

$$AbsRel = \frac{1}{N}\sum\frac{|d_t - d_p|}{d_t} \quad (5)$$

Tai pati intuityviausia klaidos metrika, parodanti vidutinę procentinę klaidą. Papildomai yra atliekama dalyba iš tikrojo atstumo ($\boldsymbol{d_t}$), taip įvertinant atstumo kontekstą (reliatyvus ar metrinis).

Jei modelio AbsRel yra 0,10, reiškia, vidutiniškai modelis klysta 10 % nuo tikrojo objekto atstumo.

RMSE (Šaknies vidutinė kvadratinė klaida):

$$RMSE = \sqrt{\frac{1}{N}\sum\left(d_t - d_p\right)^2} \tag{6}$$

RMSE naudojama norint įvertinti modelio stabilumą. Jei modelio absoliutinės santykinės klaidos reikšmė yra žema, tačiau RMSE išlieka aukšta, tai indikuoja reikšmingą klaidų pasiskirstymą. Toks metrikų neatitikimas rodo, kad nors bendras modelio tikslumas ir centrinė tendencija yra geri, tačiau modelis yra itin jautrus duomenų išskirtims arba prastai generalizuoja ribinius atvejus.

SqRel (Kvadratinė santykinė klaida):

$$SqRel = \frac{1}{N}\sum\frac{\left(d_t - d_p\right)^2}{d_t} \tag{7}$$

Metrika naudinga vertinant objektus, esančius arti kameros. SqRel identifikuoja modelio paklaidas prognozuojant artimus objektus.

Log10 (Logaritminė klaida):

$$Log10 = \frac{1}{N}\sum\left|log_{10(d_t)} - log_{10(d_p)}\right| \tag{8}$$

Metrika įvertina, kaip toli esančios objektų foninės klaidos užgožia mažų atstumų klaidas.

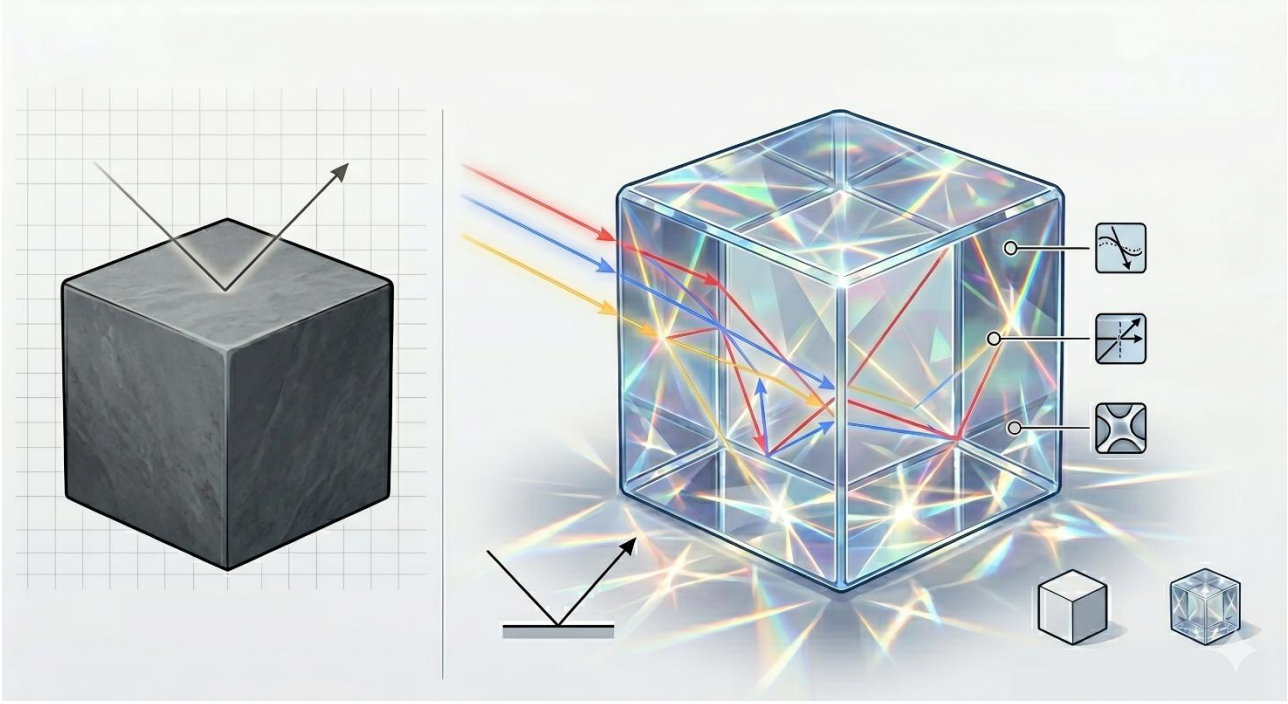
3. Duomenų sintezė

Skyriuje aprašoma gautų duomenų analizė.

3.1. Generatyvinių modelių apribojimai ir kokybės kriterijai

Vienas iš siūlomo sprendimo trūkumų yra tai, kad augmentuotų duomenų kokybė tiesiogiai priklauso nuo difuzinio modelio sugeneruotų duomenų kokybės. Priklausomybė išsiskiria į du pagrindinius elementus: sugeneruoto objekto fotorealizmą ir objekto formos išlaikymą keičiant objekto medžiagą (angl. style transfer) [43].

Modelio pasirinkimas yra kritiškai svarbus, norint gauti kokybiškus duomenis. Tačiau kai kurie užduoties aspektai yra iki šiol neišspręsti. Pakeitus objekto medžiagą iš neskaidrios į skaidrią, užtikrinti jos formą yra labai sudėtinga. Galima įvertinti ir palyginti scenos objektų kontūrus, tačiau tai užtikrina tik išorinę objekto formą, o ne vidinę daug labiau kompleksišką ir daugiamatę objekto struktūrą. Objekto forma tampa sudėtingesnė, nes užtikrinti fiziškai teisingą objekto kūno formą reikalauja teisingo šviesos sąveikos su visais objekto sluoksniais atvaizdavimo (žr. **7 pav.**).



7 pav. Objekto sąveikos su šviesa skirtumas transformuojant medžiagą į permatomą. Pabrėžiamos naujai atsiradusios refrakcijos, vidiniai atspindžiai ir kaustikos (iliustracija sugeneruota naudojantis Gemini 3 Pro Image modelį)

Modeliams neteisingai sugeneravus skaidrius objektus, gaunamas specifinis vizualus artefaktas. Nepavykęs stiklas atrodo plokščias, maža raiška, lyginant su aplinka, bei turi netolygų fokusavimo lygį, kur to paties objekto skirtingos vietos atrodo esančios skirtingu atstumu. Papildomai šie objektai praranda savo formą ir iš sudėtingų objektų tampa paprastesniais, turinčiais daug mažiau bruožų. Siekiant generuoti duomenis, kurie galėtų būti naudojami kaip MDE modelio mokymo duomenys, transformuoti objektai turi tobulai pritapti scenoje - objektas turi sukurti šešėlį ir reaguoti į skirtingus šviesos šaltinius. Atsižvelgiant į tai, kad objektas yra skaidrus, šviesa turi keliauti kiaurai objektą ir priklausomai nuo aplinkybių net būti iškreipta ar atspindėta kitoje scenos dalyje. Neteisingai

replikavus sąlygas rezultatas gali trikdyti modelį ir sukurti vizualius prieštaravimus. MDE modelis išmoktų reaguoti į neteisingas vizualines užuominas, o tai paverstų visą duomenų rinkinį nepanaudojamą treniravimo procesui.

Įdažymo užduotis yra sudėtinga ne vien dėl kūno formos išlaikymo, tačiau ir dėl naujai sukurtų scenos reikalavimų, kurie yra naujo kūno permatomos medžiagos pasekmės. Šis procesas reikalauja atitikimo daug skirtingų sąlygų atžvilgiu ir yra didžiausias apribojimo šaltinis tyrime. Visos tyrimo dalys tiesiogiai priklauso nuo įdažyto objekto kokybės, kuri yra vienareikšmiškai kontroliuojama difuzinio modelio galimybėmis.

Siekiant užtikrinti sugeneruoto objekto kokybę, įvertintos skirtingos kūno įvertinimo strategijos. Viena iš svarbiausių užduočių tyrime yra užtikrinti konvertuoto kūno atitikimą originaliam objektui. Užduotis susideda iš objekto silueto ir objekto vidinės struktūros. Objekto vidinė struktūros problema yra panaši į monokuliarinio gylio nustatymo užduoti, kuri yra korektiškai nesuformuluota, kadangi „viena 2D scenos projekcija gali turėti begalybę 3D scenų atitikmenų“. Tačiau objekto silueto užtikrinimo procesas turi daug daugiau standartinio tipo patikros opcijų.

Lengviausia ir plačiausiai naudojama strategija užtikrinti objekto siluetą naudojant difuzinio tipo generatyvinius modelius yra naudoti kaukes. Kaukė apriboja modelio veikimo zoną ir neleidžia modeliui keisti vizualių nuotraukos aspektų jei jie nepriklauso parinktai zonai. Objekto zoną galima gauti naudojantis kitu modeliu, kuris buvo treniruotas objektų aptikimo ir segmentavimo srityje. Tai pridėtų papildomą modelį į duomenų generavimo veiksmų seką, tačiau garantuotų scenos modifikacijos vietą ir objekto silueto formą. Modelio naudojimas leistų pasitelkti platesnį difuzinių modelių asortimentą, kadangi didelė įdažymo modelių dalis reikalauja objekto kaukės, o ne žodinių instrukcijų. Tačiau toks sprendimas turi kelis kardinalius trūkumus, kurie neleidžia jo naudoti tyrime. Pirmas trūkumas - pridėjus papildomą modelį į duomenų generavimo procesą susiduriama su naujomis galimomis klaidomis. Segmentavimo modelis gali neteisingai arba ne visiškai apibrėžti objektą, tai sukeltų papildomų paradoksalių scenų. Tai nėra kritinis trūkumas, nes jis galėtų būti mažiau neigiamas negu dabartinės silueto paklaidos ir pagerinti bendrą duomenų generavimo proceso rezultatą. Didžiausias strategijos ribotumas kyla iš didžiausio privalumo – zonos kontrolė. Keičiant medžiagą iš keramikos į medį pasirinkta strategija tiktų puikiai, tačiau medžiagai esant skaidriai reikalinga papildoma apimties zona, kad galima būtų pakeisti aplinkos šešėlius ir galimus atspindžius. To neatlikus susiduriama su objektais, kurie nesilaiko scenos sudarytų šviesos taisyklių ir išsiskiria nelogiškais aplinkos rėmais.

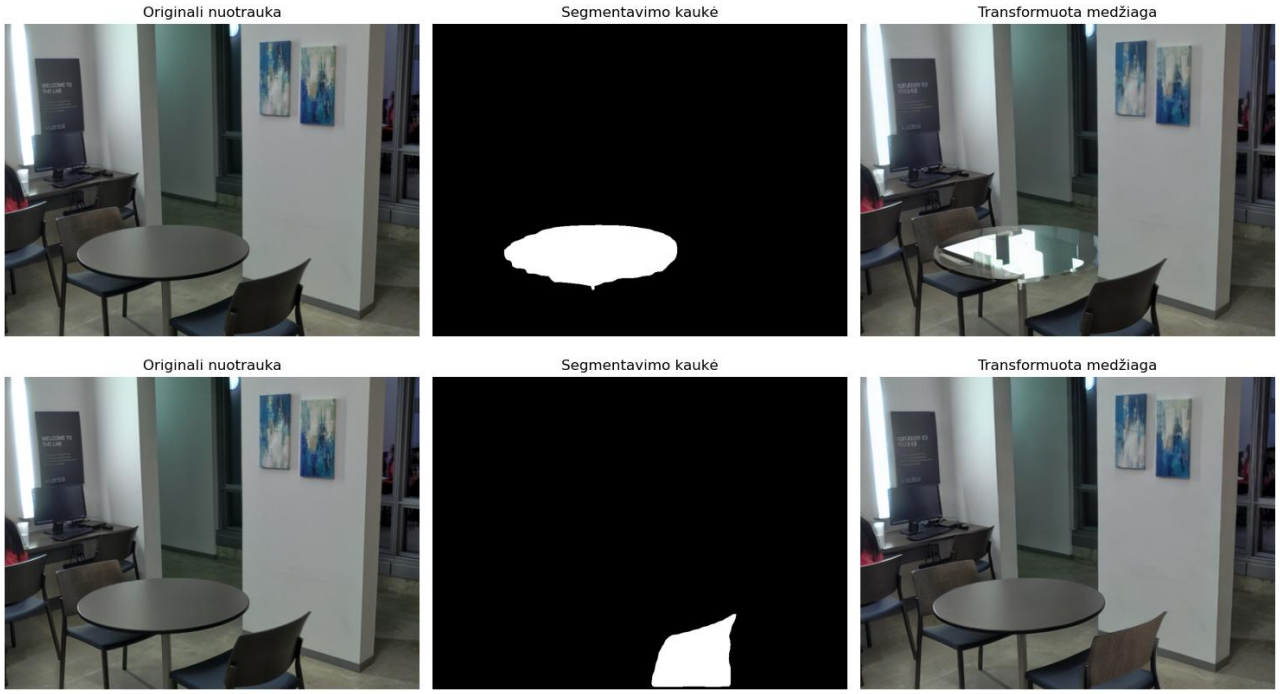
Patikrinti segmentavimo kaukės ir įdažymo procesus buvo atliekami du skirtingų modelių kombinacijų testavimo etapai. Pirmame etape buvo naudojami:

Segmentavimo modelis: maskrcnn_resnet50_fpn [44]

Difuzijos modelis: runwayml/stable-diffusion-inpainting [45]

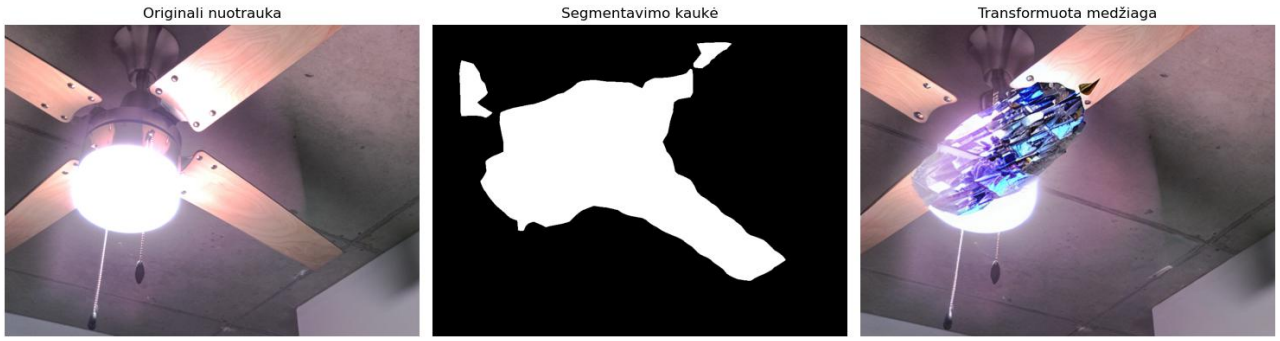
Analizuojant **8 pav.** matoma scena su stalu ir kėdėmis. Dvi eilutės nuotraukų vaizduoja vienodą sceną, tačiau skirtingus joje segmentuotus objektus. Abu pavyzdžiai pabrėžia problemą su segmentacijos modeliais. Parodytu atveju objektai segmentuoti nepilnai. Tai sudaro klaidingus duomenis atlikus difuzija, nes medžiagai pasikeitus jungtinė dalis tarp naujai sukurto stiklo ir neaptiktos stalo ar kėdės dalies atrodo nerealistiškai. Papildomai susiduriama su šešėlių ne

atitikmenimis - jei stalas yra skaidrus jo šešėlis turėtu būti minimalus, tačiau žvelgiant į abi sėdimąsias kėdžių dalis matomas stiprus šešėlis.



8 pav. SDInpainting ir ResNetFPN modelių klaidingai sugeneruotų segmentacijos kaukių ir modifikuotų scenų pavyzdžiai

Naudojant segmentacijos modelį susiduriama su papildomu apribojimu, matomu **9 pav.**. Svarbu atkreipti dėmesį į segmentacijos kaukę. Pastaroji nėra pilna, nes modelis nebuvo mokytas aptikti lubų ventiliatorius. Tai pabrėžia dar vieną segmentacijos modelių ribojimą.



9 pav. ResNetFPN modelio nesėkmingos segmentacijos kaukės pavyzdys

Žvelgiant į **10 pav.** matomi trys pavyzdžiai, kurie visi sukurią vienodą scenos paradoksą. Segmentuojant tik objektą be scenos aplinkos susiduriama su situacijomis, kur pakeitus objektą scenoje, išlieka buvusio objekto užuominos. Pirmoje nuotraukų eilutėje matomas netransformuoto šaldytuvo atspindys stalviršyje. Antroje ir trečioje eilutėse susiduriama su šešėlio prieštaravimais naujai transformuotam objektui.

10 pav. SDInpainting ir ResNetFPN modelių segmentacijos kaukės ir modifikuojamos scenos

Atliktas papildomas bandymas su naujesnėmis ir didesnėmis modelių versijomis.

Segmentavimo modelis: maskrcnn_resnet50_fpn_v2 [46]

Difuzijos modelis: diffusers/stable-diffusion-xl-1.0-inpainting-0.1 [45]

11 pav. SDXL ir ResNetFPNV2 modelių segmentacijos kaukės ir modifikuojamos scenos

Atlikus naujus bandymus susiduriama su vienodomis modelių klaidomis. **11 pav.** matomi vaizdai vizualiai transformuojami į labiau fotorealistiškus vaizdus, tačiau segmentuojami ne pilni kūnai ir transformuotų modelių pokyčiai nėra reprezentuoti scenos aplinkoje.

Dėl iškeltų reikalavimų naujai sugeneruotiems ir transformuotiems objektams užduočiai atlikti galima naudoti multimodalinius modelius. Tai yra Google Gemini ir FLUX klein tipo modeliai, kurie veikia naudojantis vaizdo ir instrukcijos duomenimis. Modeliai naudoja CLIP, transformerių architektūras, dėmesio mechanizmus ir buvo treniruoti naudojantis didelius duomenų rinkinius. Tai leidžia vartotojui pateikti modifikacijai pasirinktą vaizdą ir instrukcija, be papildomų segmentacijos ar difuzijos modelių. Multimodalinis modelis suranda objektą naudojantis pateiktos instrukcijos ir nuotraukos kombinacija. Modelis nusprendžia, kokį kiekį vaizdo modifikuoti, o tai leidžia išvengti dažnai pasikartojančios segmentacijos klaidos.

Atlikti bandymai naudojantis FLUX.2[klein] 9B versija [47]. Analizuojant **12 pav.** matomi sėkmingai įklijuoti objektai. Nors naudojama instrukcija buvo „Add a transparent object to the scene, do not change the scene keep all other aspects of the scene the same and preserve as much of the original scene as possible", tačiau sugeneruoti nauji objektai vizualiai panašiausi į plastikines plokštumas. Jie atitinka scenos iškeliamus reikalavimus ir pritampa scenoje vizualiai.

12 pav. FLUX modelio skaidraus objekto pridėjimo į sceną pavyzdžiai

Atliekant medžiagos transformacija susiduriamas su neatitikimais vaizduojamais **13 pav.**. Apie 50 % sugeneruotų vaizdų neatitiko originalios scenos. Modelis dažnai pasirenka įklijuoti naują objektą vietoj jau egzistuojančio objekto transformacijos. Transformuoti kūnai laikosi scenos ir šviesos taisyklių, tačiau didelis nesėkmingų transformacijų procentas neleidžia jo naudoti kaip transformacijas atliekančio modelio. Modifikuojant modelio hiperparametrus ir instrukciją galima pagerinti jo generuojamus rezultatus, tačiau pirminiai bandymai neatitiko lūkesčių.

13 pav. FLUX modelio medžiagos transformavimo pavyzdžiai

Atlikus bandymus buvo pasirinktas Google Gemini 3 Pro Image modelis [48]. Pastarasis kokybiškai atliko pridėjimo ir transformacijos užduotis. Detalesnė modelio rezultatų analizė aprašyta tolesniame skyriuje.

3.2. Sintetinių duomenų analizė

Analizuojant pateiktus vaizdus, pavyzdžių išdėstymas seka ta pačia struktūra, kaip parodyta **14 pav.**.

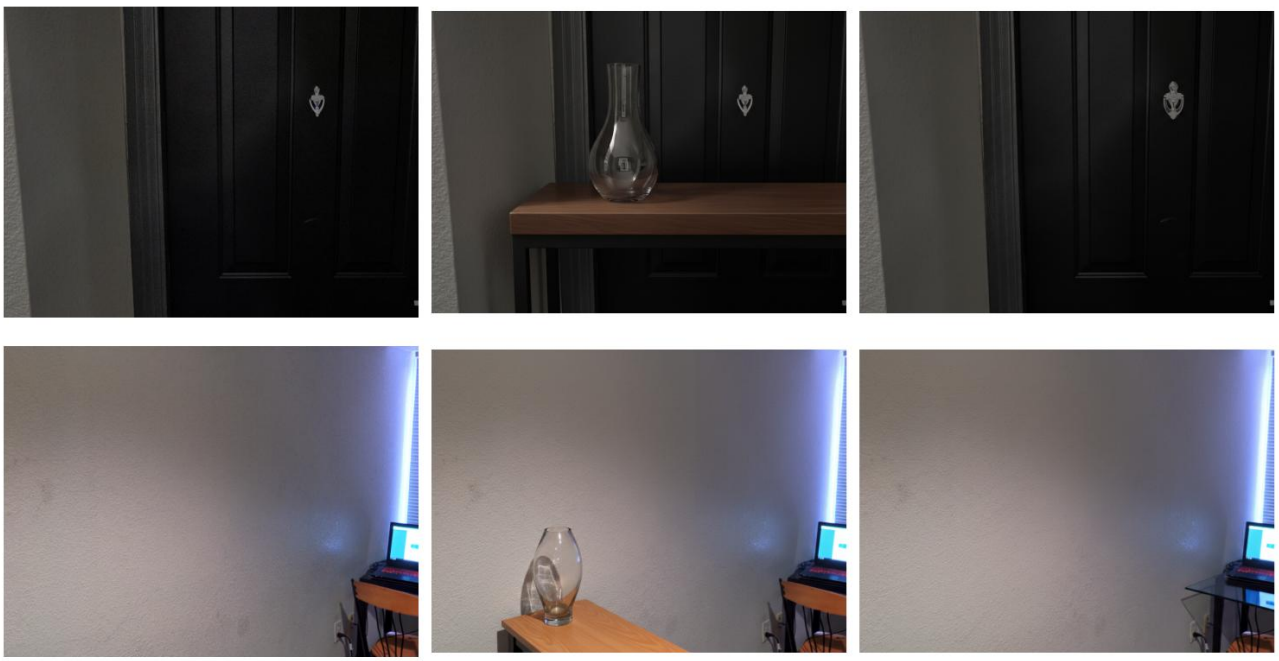
Modelio hiperparametrai nėra laisvai prieinami vartotojui, taip siekiant supaprastinti vartojimą. Modelis priima techninius sprendimus remiantis vartotojo parašyta žodine instrukcija.

14 pav. Scenų, kuriuose generavimo procesas atliktas sėkmingai, pavyzdžiai

Nagrinėjant **14 pav.**, demonstruojami trys skirtingi scenų scenarijai. Į scenas įterpti sintetiniai objektai – tai trys ore sklandančios vazos. Tokios naujos scenos nėra dažnai pasitaikančios ir jas gana sunku atkurti realybėje, tačiau vizualiai jos gerai prisitaiko prie scenos apšvietimo. Visi šie objektai turi šešėlį ir išlaiko scenos informaciją, kuri originalioje nuotraukoje buvo matoma, bet dabar yra uždengta nelambertinio kūno.

Dvi iš trijų transformuotų scenų patyrė nedidelį geometrinio identiteto praradimą, o naujai sugeneruoti objektai ne visiškai tiksliai atspindėjo savo realius atitikmenis. Vaizde *00000_00003_indoors_170_000.png* (pirmoje eilutėje) metalinis elektros komponentų dangtelis paverčiamas stiklo plokštuma. Ji nevisiškai atitinka originalaus objekto geometrinius apribojimus, praranda nedidelius įdubimus, bet išlaiko siluetą ir bendrą reliatyvią vietą scenoje. Tai sukels neatitikimą tarp tikrovės ir vizualinių duomenų.

Didesnė problema matoma **15 pav.**: pridedant į sceną naują, joje neegzistavusį objektą, modelis be atskiro nurodymo prideda ir papildomą palaikantį objektą, aprašytu atveju – stalą, skirtą vazai paremti ar pastatyti. Tai nutinka maždaug 2 % atvejų, kai modeliui pateikiama lygiai ta pati užklausa. Problemą galima spręsti koreguojant modelio hiperparametrus, arba taip pat įmanoma naujus priedus išfiltruoti, kadangi jie įneša į sceną radikaliai kitokias spalvas ir pakeičia scenos spalvų paletę pastebimai labiau nei tai padarytų vien tik skaidrus objektas.



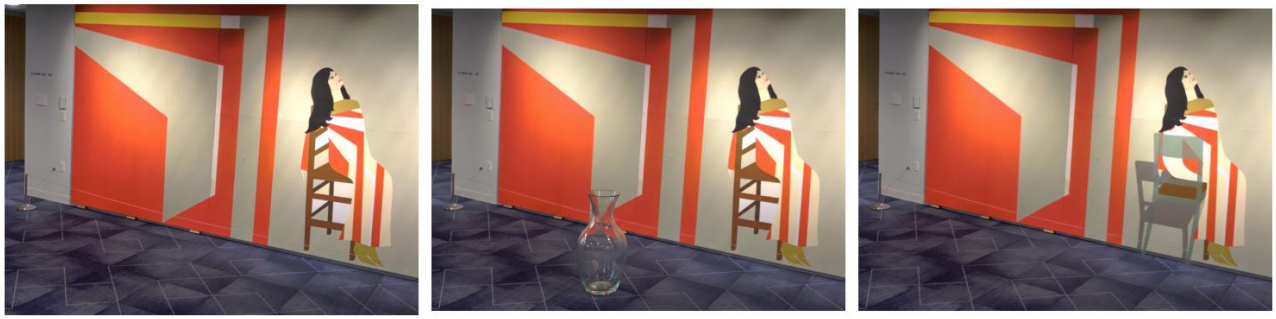
15 pav. Scenos pavyzdžiai, kuriuose pridedant naują skaidrų kūną kartu pridedamas ir neskaidrus objektas

Tuo tarpu **16 pav.** pavaizduoti didesni geometriniai neatitikimai. Modelis vienu atveju stiklu padengia pusę scenos ir ją suplokština (viršutinė eilutė), arba sukuria įdubimus bei nenatūraliai pasuka objektą (apatinė eilutė). Automatizuotas vaizdų tikrinimas reikalautų papildomos intelektualios sistemos, galinčios įvertinti jau ir taip sudėtingų nelambertinių objektų vientisumą ir nuoseklumą.

16 pav. Medžiagos transformacijos pavyzdžiai, kuriuose bendra scenos geometrija išlaikyta, tačiau sukuriama transformuoto kūno geometrijos klaida

Galiausiai, **17 pav.** išryškėja problema, kylanti dėl objektų trūkumo scenoje. Gavus nurodymą pakeisti scenoje dominuojantį objektą ir nesugebėjus aptikti tinkamo pirmame plane esančio kūno,

modelis atlieka nenumatytas transformacijas, pavyzdžiui, pakeičia 2D sienos tapybos elementų medžiagą. Sienos paveiksle pavaizduota ant kėdės sėdinti moteris, todėl modelis kėdės medžiagą pakeičia į skaidrią, stiklą primenančią medžiagą. Kadangi nepavyksta transformuoti realaus, fizinio scenos objekto, toks atvejis laikomas nesėkmingu.



17 pav. Klaidingai konvertuoto kūno scenos pavyzdys. Sienos dekoro objekto transformacija

Siekiant išvengti situacijų, kuriuose susidūrus su scena be objektų sugeneruojama klaidingai, buvo modifikuota transformavimo instrukcija. Naujoje instrukcijoje aprašomas scenarijus kuriame nėra aptinkamas scenos objektas. Taip įvykus modifikuojamos scenos sienos ir grindys ar lubos.

Sena instrukcija:

„Identify a prominent furniture piece. Change its material to transparent glass. Preserve shape, shadow, position. Do not change background."

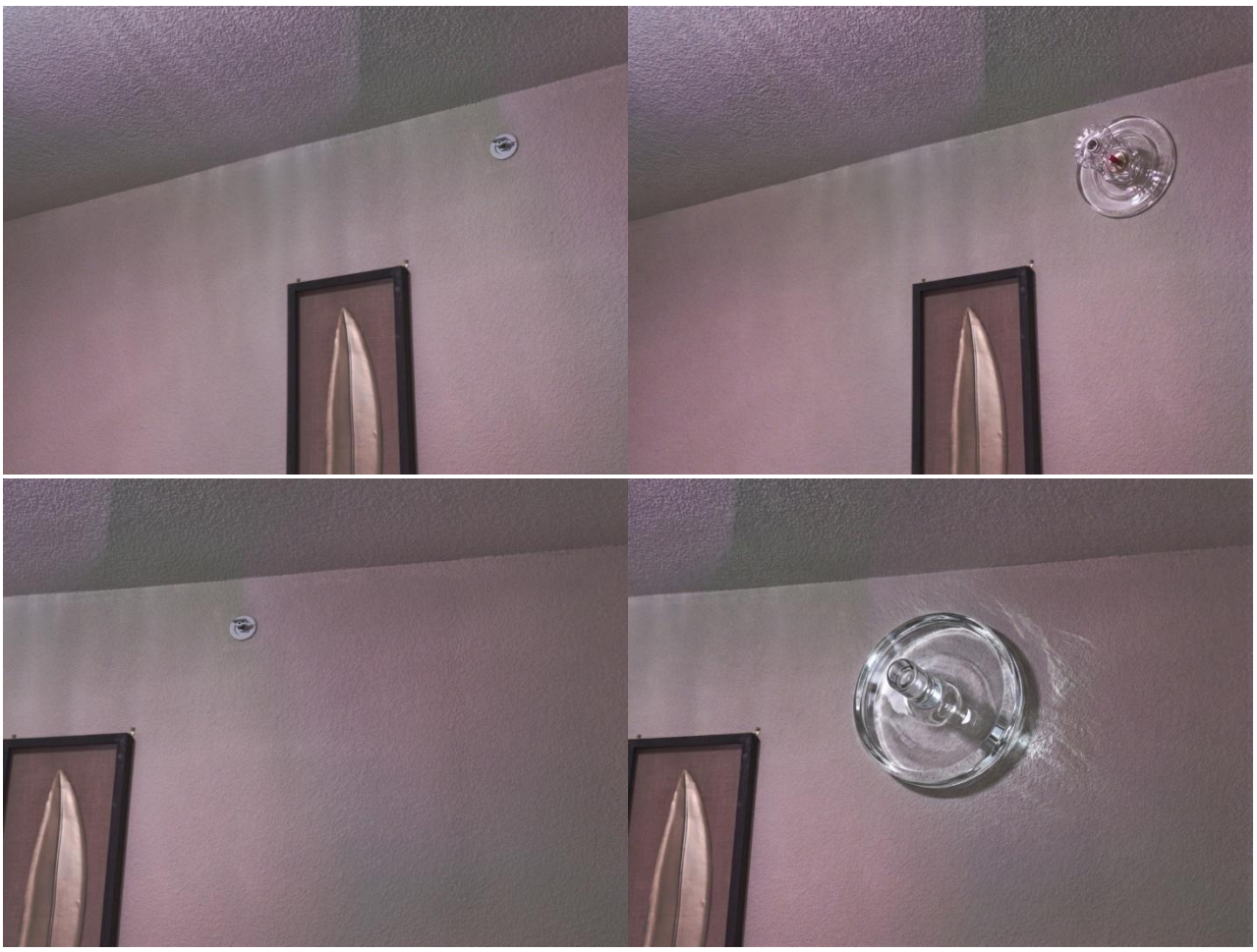
Nauja instrukcija:

„Change the material of an existing object or a small part solid structural surface (like a wall or floor segment) into transparent glass. You must perfectly preserve its original 3D shape, volume, and boundaries. Do not introduce any structural irregularities. The scene's geometry must remain untouched. Do not add new objects only convert already existing ones."

Nauja instrukcija suteikia modeliui galimybę prisitaikyti prie nuotraukos reikalavimų ir sukurti scenas su skaidriais objektais, net scenoms esant tuščioms. Kaip matoma **18 pav.** naujai sukuriamos scenos turi skaidrių objektų ir išlaiko originalią scenos gylio struktūrą.

18 pav. Medžiagos transformavimo užduoties įvykdymas scenai neturint aptinkamų objektų

Naudojantis nauja instrukcija išvengiama tuščių nemodifikuotų scenų, tačiau instrukcijos prioritetas išlieka objektams egzistuojančioms scenoje. Tai sukelia papildomų problemų, ypač pastebimų kai modelio pasirinktas modifikavimui objektas yra labai mažas (žr. **19 pav.**). Modeliui pasirinkus modifikuoti per mažą objektą, susiduriama su transformacijos klaida, kurios metu objekto forma ir dydis yra radikaliai pakeičiami. Gautas rezultatas yra neteisingas ir sukelia didelį ne atitikmenį su scenai priklausančiu gylio žemėlapiu.



19 pav. Medžiagos transformavimo klaida susiduriant su mažais objektais

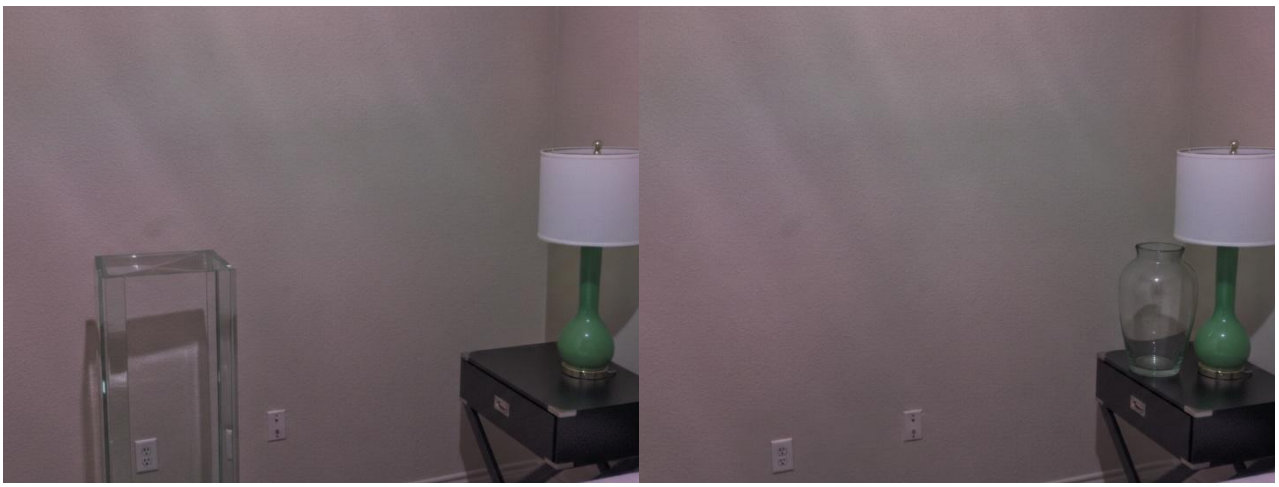
Analizuojant pavyzdžius matyti, kad skaidrių duomenų sintezė yra įmanoma ir kai kuriais atvejais jau teikia tenkinančius rezultatus. Tačiau kitais atvejais modeliui nepavyksta išlaikyti scenos struktūros ir nauji objektai yra priešingi tikrojo gylio žemėlapiams.

3.2.1. Scenų variacija

Generuojant naujus objektus scenoje galima naudoti tapačią bazinę sceną ir gauti daug skirtingų scenos variacijų. Tai leidžia naudoti labai mažos apimties duomenų rinkinius ir sintetiškai juos praplėsti. Toks pritaikymo principas leidžia modelį mokyti ir validuoti naudojantis ta pačia bazine scena.

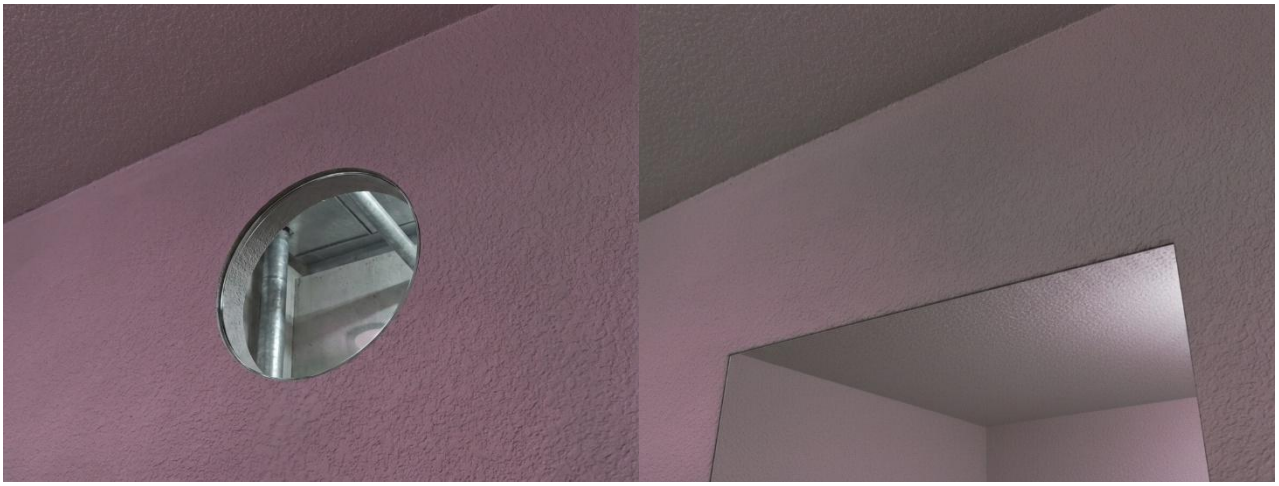
Metodas pritaikomas medžiagos konvertavimo duomenų ir naujo skaidraus objekto įklijavimo duomenų rinkiniams. Teksto instrukcijomis valdomas difuzijos modelis leidžia naudoti pakankamai abstrakčias instrukcijas, leidžiančias modeliui prisitaikyti prie skirtingų scenų ir gauti skirtingą rezultatą naudojant vienodą pradinę sceną.

Generuojant nuotraukas fono aptikimui, naujai įklijuotas objektas gali būti bet kurioje scenos vietoje (žr. **20 pav.**). Naujai sugeneruoti objektai gali nesilaikyti gravitacijos dėsnių ir tiesiog būti viduryje scenos, be jokios papildomos struktūros laikančios objektą. Tai leidžia sukurti daug įvairių scenos variacijų naudojantis vienoda modelio instrukcija.



20 pav. Skaidraus objekto įklijavimo scenos variacijos pavyzdys

Transformuojant jau egzistuojančius objektus į skaidrius ar atspindinčius paviršius susiduriama su didesniais abstrakčios instrukcijos ribojimais. Siekiant transformuoti aiškiai scenoje figūruojantį objektą ir vengti mažų vos pastebimų scenos detalių modifikavimo, instrukcija turi būti aiški ir įvardinti prioritetą jau egzistuojančioms scenos detalėms. Tačiau instrukcijai aprašant tik kambario baldą ar duris, nėra įvertinami visi galimi scenarijai. Jei nuotraukoje matoma tik sienos dalis, be jokių papildomų baldų ar kitų lengvai transformuojamų objektų, modelis priverstas ignoruoti instrukciją ir grąžiną neteisingai modifikuotą nuotrauką arba nemodifikuotą sceną. Instrukciją galima praplėsti įtraukiant alternatyvą neaptikus lengvai atpažįstamų didelių objektų scenoje. Alternatyvi instrukcija nurodo modeliui konvertuoti mažą plotą jau egzistuojančios sienos ar lubų. Visos kitos instrukcijos apie scenos geometrijos išlaikymą lieka nepakeistos.



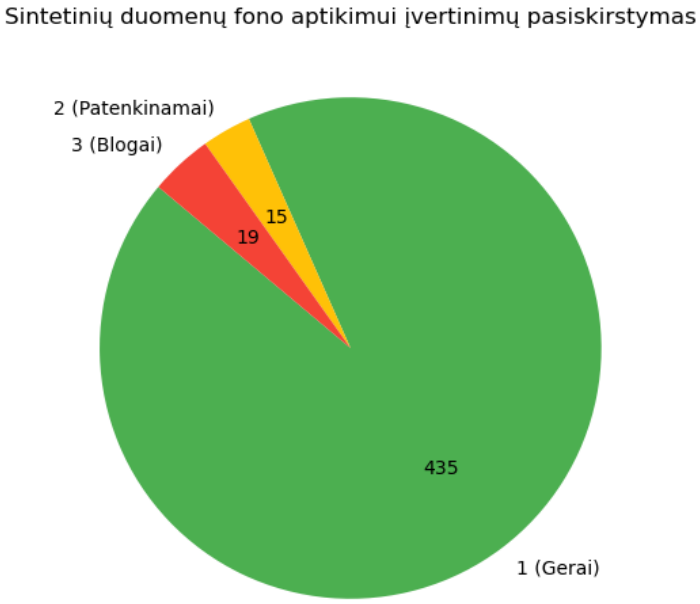
21 pav. Medžiagos transformavimo scenos variacijos pavyzdys

Naujos scenos variacijos (žr. **21 pav.**) vizualiai skiriasi viena nuo kitos ir gali būti naudojamos, kaip validacijos rinkinio dalis. Modeliui išmokus didžiąją dalį scenos iš mokymo rinkinio, tenka susidurti su vizualiai skirtinga scenos variacija, kuri pabrėžia norimo veikimo principą. Skirtingų nuotraukų variacijų egzistavimas skatiną modelį pritaikyti vienodą logiką susiduriant su skirtingais objektais.

3.3. Duomenų analizės rezultatai

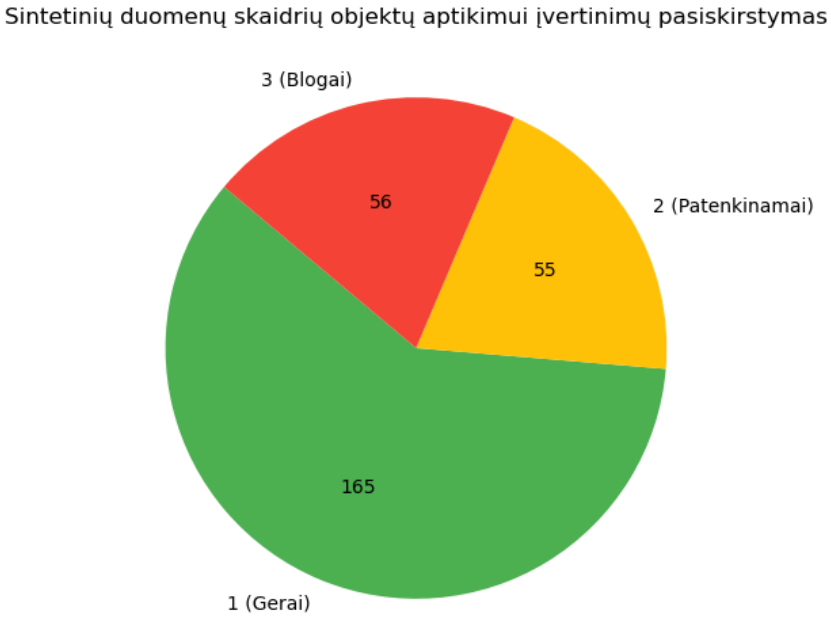
Dėl scenų įvairovės ir modifikavimo tipo, siekiant įvertinti sugeneruotus duomenis buvo pasitelkiamas rankinis duomenų tikrinimas. Vertintojui yra suteikiami trys vertinimo lygiai. Lygiai bendrai apibūdinami taip:

1. Gerai: vizualiai duomenys tobulai atitinka sceną.
2. Patenkinamai: pridedami minimalūs vizualiniai papildai (stiklai prilaikantys kabeliai ar minimalus formos pokytis). Patenkinami duomenys gali būti naudojami mokymui, tačiau pakels klaidų kiekį dėl savo kokybės.
3. Blogai: duomenys neturėtu būti naudojami. Kardinaliai pakeista scena (pridedami nauji scenoje neegzistuojantys kūnai arba neišlaikoma scenos geometrinė forma).



22 pav. Fono aptikimui sugeneruotų duomenų įvertinimų pasiskirstymo skritulinė diagrama

Vertinant **22 pav.** vaizduojamą skritulinę diagrama pastebima, kad tik 7,2 % fono modelio duomenų yra įtartini. Tai reiškia, kad sugeneruotos scenos yra didžiąja dalimi naudojamos ir sukelia mažą kiekį ne atitikmenų tarp gylio žemėlapių ir vizualinių duomenų.



23 pav. Pirmo plano aptikimui sugeneruotų duomenų įvertinimų pasiskirstymo skritulinė diagrama

Analizuojant **23 pav.** vaizduojamą skritulinę diagramą, regimas ženkliai didesnis neteisingų duomenų kiekis. Net 40,2 % visų sugeneruotų duomenų yra vizualiai neatitinkantys scenoje ir sukeliantys didelį skirtumą tarp gylio žemėlapių ir vizualinių duomenų.

Apibendrinant, sugeneruotų duomenų kokybė koreliuoja su užduoties sudėtingumu. Naujų objektų pridėjimas į sceną dažniausiai atliekamas sėkmingai, o medžiagos konvertavimas sukelia daug ne atitikimų su gylio žemėlapiais ir turi prastą vidutinę duomenų kokybę.

4. Monokuliarinio gylio modelių eksperimentiniai rezultatai

Skyriuje aprašoma projekto eksperimentai ir rezultatai. Skyrių sudaro pagrindiniai poskyriai aprašantys eksperimentų aplinką ir sąlygas, duomenų rinkinio statistiką, eksperimentines mokymo dalis, rezultatų analizę.

4.1. Eksperimentų aplinka ir sąlygos: techninė įranga programinė įranga

4.1.1. Techninė įranga

2 lentelėje pateikta techninė įranga naudota atlikti projektą ir realizuoti eksperimentinio tipo mokymo dalis.

2 lentelė. Asmeninio kompiuterio sudedamosios dalys

Vaizdo plokštė (GPU):	NVIDIA GeForce RTX 4060ti (16 GB vidinė atmintis VRAM)
Procesorius (CPU):	Processor 13th Gen Intel(R) Core(TM) i7-13700F, 2100 Mhz, 16 Core(s), 24 Logical Processor(s)
Operatyvioji atmintis (RAM):	32 GB
Atminties tipas (Diskas):	Lexar SSD NM620 NVMe 2TB

Eksperimentinės mokymo dalys ir difuzinių modelių testavimas buvo atliktas naudojantis asmeniniu kompiuteriu. Tačiau dėl pasirinkto Google Gemini Image 3 difuzinio modelio, treniravime naudotos modifikuotos sintetinės nuotraukos buvo generuojamos naudojantis Google API debesies paslaugomis. Techninės įrangos specifikacijos, kurias naudoja Google nėra laisvai prieinamos, todėl vienintelis techninės įrangos įvertis yra nuotraukos išsiuntimo, generavimo, gavimo laiko trukmės suma.

- Google debesies paslaugų vienos nuotraukos generavimo bendra laiko trukmė (sekundėmis): ~19 s
- Vienos nuotraukos generavimo kaina (įskaičiuojant instrukcijos ir vaizdo generavimo žetonai (angl. tokens): apie 0,134 $ per nuotrauką

Atlikus pagrindinius generavimo procesus galutinė kaina už Google API paslaugas pasiekė 19,57 €. Į šią kainą įskaičiuota sėkmingai ir nesėkmingai sugeneruotų vaizdų kainų suma (16,17 €) ir PVM mokestis (3,40 €). Taip pat patirta papildomų, tiksliai neįvertintų išlaidų konvertuojant valiutą iš JAV dolerių į eurus.

4.1.2. Programinė įranga

3 lentelėje pateikta darbe naudota programinė įranga.

3 lentelė. Programinė įranga

Operacinė sistema (OS):	Microsoft Windows 11 Pro for Workstations
Programavimo kalba ir versija:	Python 3.11.9
Papildomos bibliotekos (pagrindinės):	
PyTorch:	2.9.1+cu126
CUDA:	12.6
Torchvision:	0.24.1+cu126
NumPy:	2.3.1
Pillow:	10.4.0
OpenCV:	4.11.0
Transformers:	5.2.0
Google GenAI:	1.57.0
MathPlotLib:	3.9.3
Generatyvinio dirbtinio intelekto įrankis:	Google Gemini API

Priklausomai nuo eksperimentinių reikmių, buvo naudojamos skirtingos bibliotekų versijos siekiant geriausiai įvertinti alternatyvius duomenų generavimo ir treniravimo metodus. Išvardintos papildomų bibliotekų versijos apibūdiną galutinę treniravimo aplinką naudojantis DepthAnythingV2 ir Google Gemini API.

4.2. Duomenų rinkinio statistika

Eksperimentiniams tyrimams atlikti buvo paruoštas specializuotas duomenų rinkinys, suformuotas originalaus „DIODE“ (angl. Dense Indoor and Outdoor Depth Dataset) duomenų rinkinio pagrindu. Siekiant priversti modelį atpažinti stiklo ir atspindžių paviršių savybes bei suprasti geometriją esančią už jų, tyrimo duomenys buvo padalinti į dvi tikslines grupes: fono (angl. background) ir pirmojo plano (angl. foreground) vaizdų rinkinius.

Atliekant mokymą duomenų rinkiniai buvo skirstomi naudojantis 80/20 logika. Aštuoniasdešimt procentų viso duomenų rinkinio buvo naudojama mokymui, o dvidešimt procentų validavimui.

4.2.1. Pirmojo plano duomenų rinkinys (angl. Foreground dataset)

Rinkinį sudaro vaizdai, kuriuose objektai buvo konvertuoti į skaidrius ir atspindinčius paviršius naudojantis generatyvinio dirbtinio intelekto metodus. Rinkinio paskirtis išmokyti modelį aptikti skaidrius ir atspindžius paviršius lygiai taip pat, kaip ir bet kokį kitą objektą scenoje.

- Rinkinio apimtis: 276 modifikuotų vaizdų.

- Specifika: Kiekvienam RGB vaizdui priskirti du papildomi failai, originalus gylio žemėlapis (.npy formatu) reprezentuojantis tikrąjį atstumą iki nemodifikuotų objektų, gylio žemėlapių išskirčių kaukė (.npy formatu), nurodanti neteisingai nuskaitytas gylio žemėlapio vertes.

4.2.2. Fono duomenų rinkinys (angl. Background dataset)

Fono duomenų rinkinį sudaro vaizdai, kurie modifikuoti įpaišant naujus, originalioje scenoje neegzistuojančius objektus naudojantis generatyvinio dirbtinio intelekto metodais. Toks duomenų rinkinys priverčia besimokantį modelį ignoruoti skaidrius vaizdus ir kuo tiksliau atkurti scenos geometriją, kuri egzistuoja ir yra dalinai matoma už skaidraus objekto.

- Rinkinio apimtis: 469 modifikuotų vaizdų.

- Specifika: Kiekvienam RGB vaizdui priskirti du papildomi failai, originalus gylio žemėlapis (.npy formatu) reprezentuojantis tikrąjį atstumą iki nemodifikuotų (naujų) objektų, gylio žemėlapių išskirčių kaukė (.npy formatu), nurodanti neteisingai nuskaitytas gylio žemėlapio vertes.

4.3. Mokymas: eiga, mokymosi kreivės ir rezultatų aptarimas

4.3.1. Mokymo dinamika

Mokymo metu naudotas jau ištreniruotas bazinis DepthAnythingV2 modelis. Dėl pasirinkimo modelius mokyti lokaliai, nuspręsta treniruoti mažiausio dydžio „S“ modelių variacijas. Kadangi mokymas pradedamas modeliui jau turint iš anksto nustatytus svorius, tikimasi, kad modelis mokysis labai greitai ir klaida (angl. loss) kris sparčiausiai pirmas kelias mokymo epochas.

Atsižvelgiant į mažą mokymo duomenų kiekį, pasirinkta 50 epochų maksimali riba. Papildomai, siekiant išvengti modelių persimokymo ir sumažinti skaičiavimo resursų naudojimą, pasirinkta naudoti ankstyvą mokymosi sustabdymą. Validacijos kreivei nepasiekus naujo minimumo per 10 epochų mokymas buvo stabdomas.

4.3.2. Duomenų kiekio įtaka

Tyrimui atlikti naudoti 745 sintetiniai vaizdai (duomenų paskirstymas pateiktas **4** ir **5 lentelėse**). Siekiant įvertinti duomenų įtaką modelio gebėjimui išmokti matyti stiklą ir už jo esančią geometriją, duomenys buvo naudojami etapais. Mokymas atliktas kelis kartus pasitelkiant vis daugiau duomenų. Pasirinktas duomenų didinimo žingsnis buvo 100 nuotraukų.

4 lentelė. Fono aptikimo modelių duomenų pasiskirstymas

Modelio pavadinimas	Duomenų kiekis (nuotraukos)	Mokymo duomenys	Validacijos duomenys	
			Matyti	Nematyti
model_100_nuotrauku	193	100	0	93
model_200_nuotrauku	293	200	0	93
model_300_nuotrauku	393	300	0	93
model_Visi_duomenys	469	376	0	93
model_100_nuotrauku_dub	193	100	93	0
model_200_nuotrauku_dub	293	200	93	0
model_300_nuotrauku_dub	393	300	93	0
model_Visi_duomenys_dub	469	376	93	0
model_100_nuotrauku_50_50	193	100	46	47
model_200_nuotrauku_50_50	293	200	46	47
model_300_nuotrauku_50_50	393	300	46	47
model_Visi_duomenys_50_50	469	376	46	47

Fono aptikimui buvo mokyti 12 modelių. Pagrindiniai skirtumai tarp modelių pasireiškia mokymo aibės dydžiu ir validacijos aibės struktūra.

5 lentelė. Pirmo plano aptikimo modelių duomenų pasiskirstymas

Modelio pavadinimas	Duomenų kiekis (nuotraukos)	Mokymo duomenys	Validacijos duomenys	
			Matyti	Nematyti
model_100_nuotrauku_fg	155	100	0	55
model_200_nuotrauku_fg	255	200	0	55
model_Visi_duomenys_fg	276	221	0	55
model_100_nuotrauku_dub_fg	155	100	55	0
model_200_nuotrauku_dub_fg	255	200	55	0
model_Visi_duomenys_dub_fg	276	221	55	0
model_100_nuotrauku_50_50_fg	155	100	27	28
model_200_nuotrauku_50_50_fg	255	200	27	28
model_Visi_duomenys_50_50_fg	276	221	27	28

Pirmojo plano aptikimui mokyti 9 modeliai. Dėl mažesnio duomenų kiekio, skirtumas tarp modelio, mokyto naudojantis 200 nuotraukų, ir visus galimus duomenis skiriasi tik 21 nuotrauka.

4.3.3. Mokymosi kreivės: fono gylio žemėlapių modeliai

1) Matyti duomenys (_dub modeliai)

Naudojantis validacijos rinkiniu, kuris yra sudarytas vien tik iš scenų variacijų, kurios taipogi yra naudojamos mokymo duomenų rinkinyje, gaunami validacijos klaidų įverčiai yra ženkliai geresni negu visų kitų naudotų strategijų (žr. **24 pav.**). Matomos validacijos kreivės beveik tobulai seka mokymo kreives. Tai pabrėžia didžiausią problemą su minima strategija – duomenų persimokymą

(angl. overfitting). Kadangi naujas sintetinis objektas yra maža scenos dalis, modeliui „mintinai išmokus“ didžiąją dalį scenos, teisingai prognozuojama didžioji dalis scenos gylio. Žvelgiant į mokymo kreives matoma, kad tam tikrose vietose validacijos kreivė kertasi su mokymo kreive ir netgi ją aplenkia (žiūrėti **24 pav.**, trečią grafiką).

2) Nematyti duomenys (baziniai modeliai)

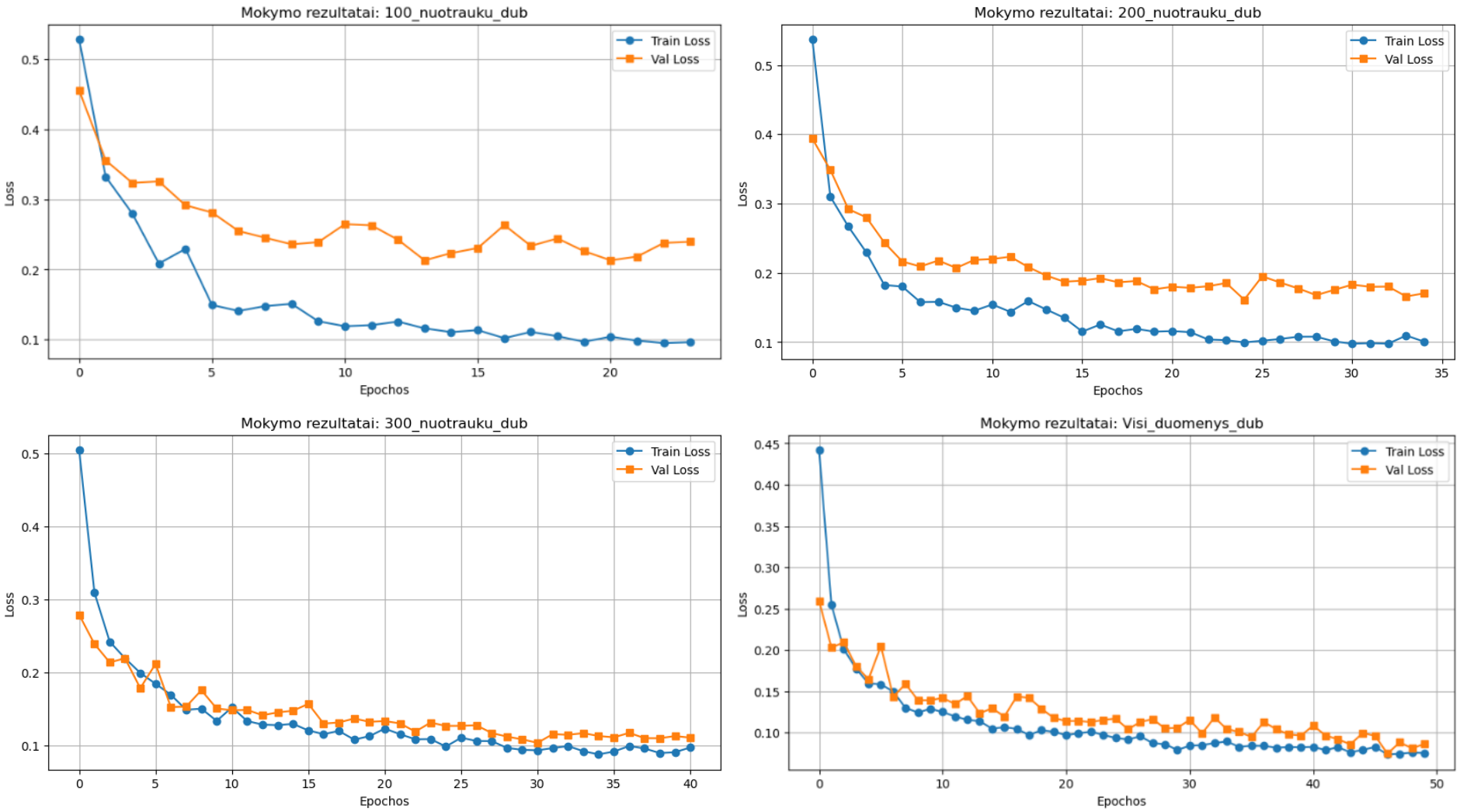
Naudojantis pilnai nematytą validacijos duomenų rinkinį pasiekiamas standartinis modelio mokymo scenarijus. Mokymosi kreivės tokiu atveju yra labiau atitrūkusios viena nuo kitos ir duomenų kiekiui didėjant tarpas mažėja. Tai yra labiausiai standartinė modelio mokymo praktika. Validacijos rinkinys leidžia simuliuoti modelio gebėjimą sudaryti gylio žemėlapius naudojantis DIODE duomenų rinkinį surinkusia kamera.

25 pav. matyti, kad validacijos kreivė nepradeda kilti, kas indikuoja modelio nepersimokymą. Net naudojant 100 nuotraukų mokymo duomenų rinkinį, validacijos kreivė išlieka horizontali ir mokymo procesas sustabdomas po 16 epochų išvengiant persimokymo. Modelis mokytas naudojant 100 nuotraukų geriausią rezultatą pasiekė šeštoje epochoje. Tai indikuoja, kad modeliui trūksta mokymo duomenų.

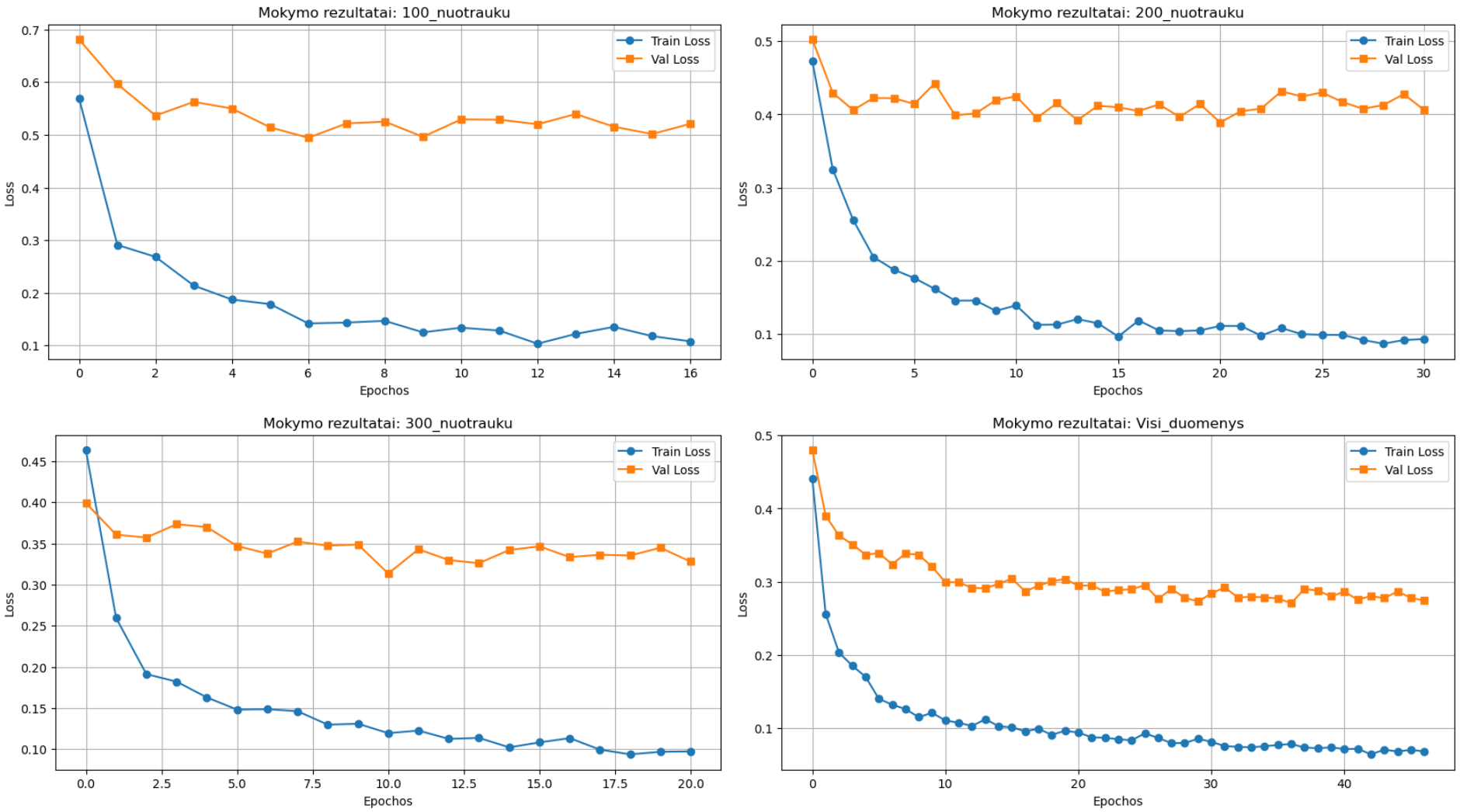
1) Hibridinė sistema (_50_50 modeliai)

Hibridinė duomenų maišymo sistema leidžia naudoti scenų variacijas validacijos duomenų rinkinyje. Tai dirbtinai pakelia gautus validacijos rezultatus.

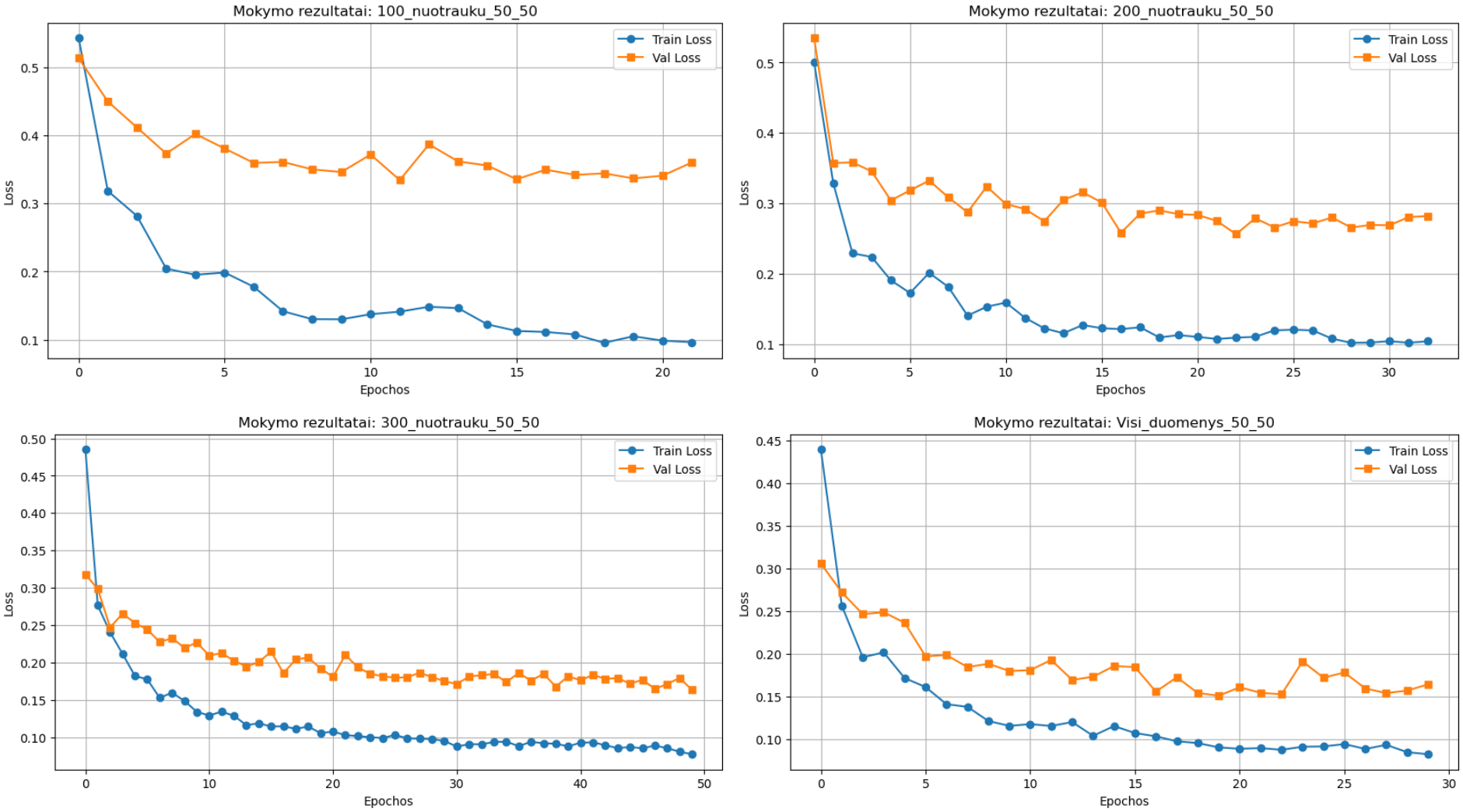
26 pav. pavaizduotos hibridinės (50 % matytų, 50 % nematytų scenų) strategijos mokymosi kreivės. Kadangi tyrimo metu buvo taikomas tikslinamasis mokymas (angl. fine-tuning), visuose modeliuose stebimas itin greitas klaidų įverčio kritimas. Didžioji dalis pokyčių įvyksta pirmuose 5 epochose. Tai atitinka lūkesčius (greitas prisitaikymas prie duomenų rinkinio), kadangi modelio svoriai yra iš anksto nustatyti ir modeliams reikia prisitaikyti tik prie kameros specifikacijos ir naujų sintetinių skaidrių kūnų logikos.



24 pav. Matytų duomenų strategijos fono aptikimo modelių mokymosi kreivė



25 pav. Nematytų duomenų strategijos fono aptikimo modelių mokymosi kreivės



26 pav. Hibridinės strategijos fono aptikimo modelių mokymosi kreivės

Analizuojant 100 nuotraukų modelį, treniravimo klaida leidžiasi iki ~0,1, tačiau validacija jau po maždaug 8-10 epochų stabilizuojasi ir pradeda lengvai banguoti ar net kilti

4.3.4. Mokymosi kreivės: pirmo plano gylio žemėlapių modeliai

Skyriuje aprašomi pirmo plano gylio žemėlapių sudarymo modelių treniravimo mokymo eiga. Įvertinamos skirtingos mokymo strategijos ir duomenų stokos įtaka modelio mokymui.

Dėl mažesnio duomenų kiekio, kiekviena primojo plano modelių mokymo duomenų rinkinių aibė turėjo mažesnį kiekį etapų.

3) Matyti duomenys (_dub modeliai)

27 pav. matomos skirtingų duomenų aibių mokymosi ir validacijos kreivės. Lyginant jas su fono modeliais matomas stiprus skirtumas. Didžiausias skirtumas pastebimas naudojantis 100 nuotraukų mokymo aibę. Fono modelis pasiekė apie ~0,2 validacijos klaidos įvertį, o pirmojo plano modelis tiktai ~0,32. Svarbu paminėti, kad skiriasi ne vien mokymo aibės struktūra, bet ir validacijos aibė. Fono validacijos aibę sudaro 93 nuotraukos, o pirmojo plano 55, daro įtaką validacijos aibės įverčių skirtumui. Tuo pačiu, dar didesnis skirtumo faktorius yra sintetinių duomenų kokybė. Kaip minėta ankščiau, medžiagos transformavimo užduotis yra ženkliai sudėtingesnė, kas atsispindi sugeneruotų duomenų kokybėje. Modeliui susiduriant su priešingybėmis mokymo rinkinyje kyla klaidų tikimybė.

Mokymo duomenų aibei padidėjus, kreivės supanašėja į fono modelių kreives. Vietomis validacijos kreivės aplenkia mokymo kreives ir atstumas tarp jų išlieka mažas viso mokymo metu. Tai pasiekiama naudojantis jau matytas modeliui scenas ir nėra geras modelio kokybės įvertinimo rodiklis. Dėl didelio validacijos ir mokymo duomenų panašumo validacijos kreivės yra iškraipomos ir neatspindi realaus pasaulio scenarijų.

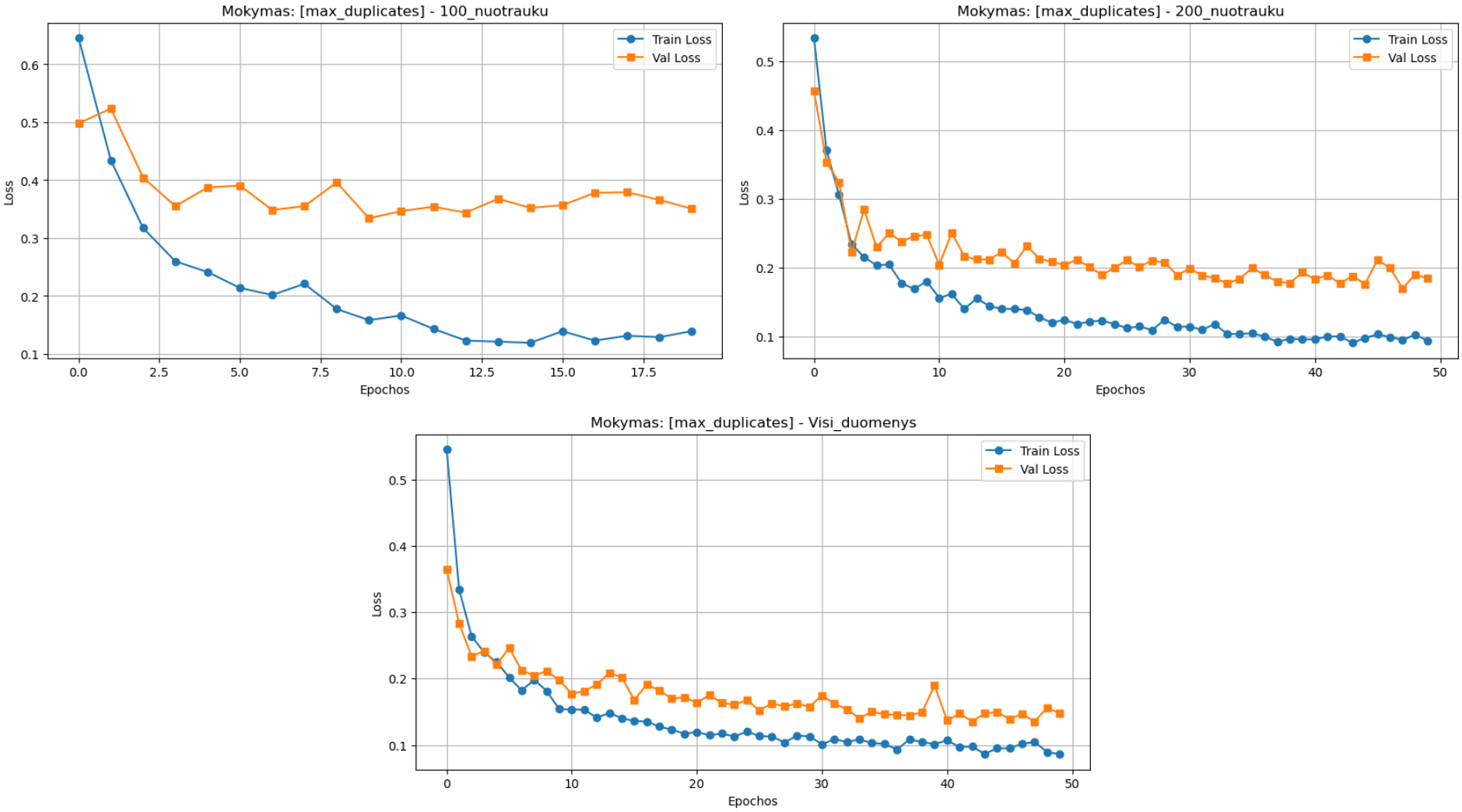
Tai reikalauja papildomo testavimo naudojantis duomenų rinkiniais, kurie nėra struktūriškai identiški mokymo duomenims.

4) Nematyti duomenys (baziniai modeliai)

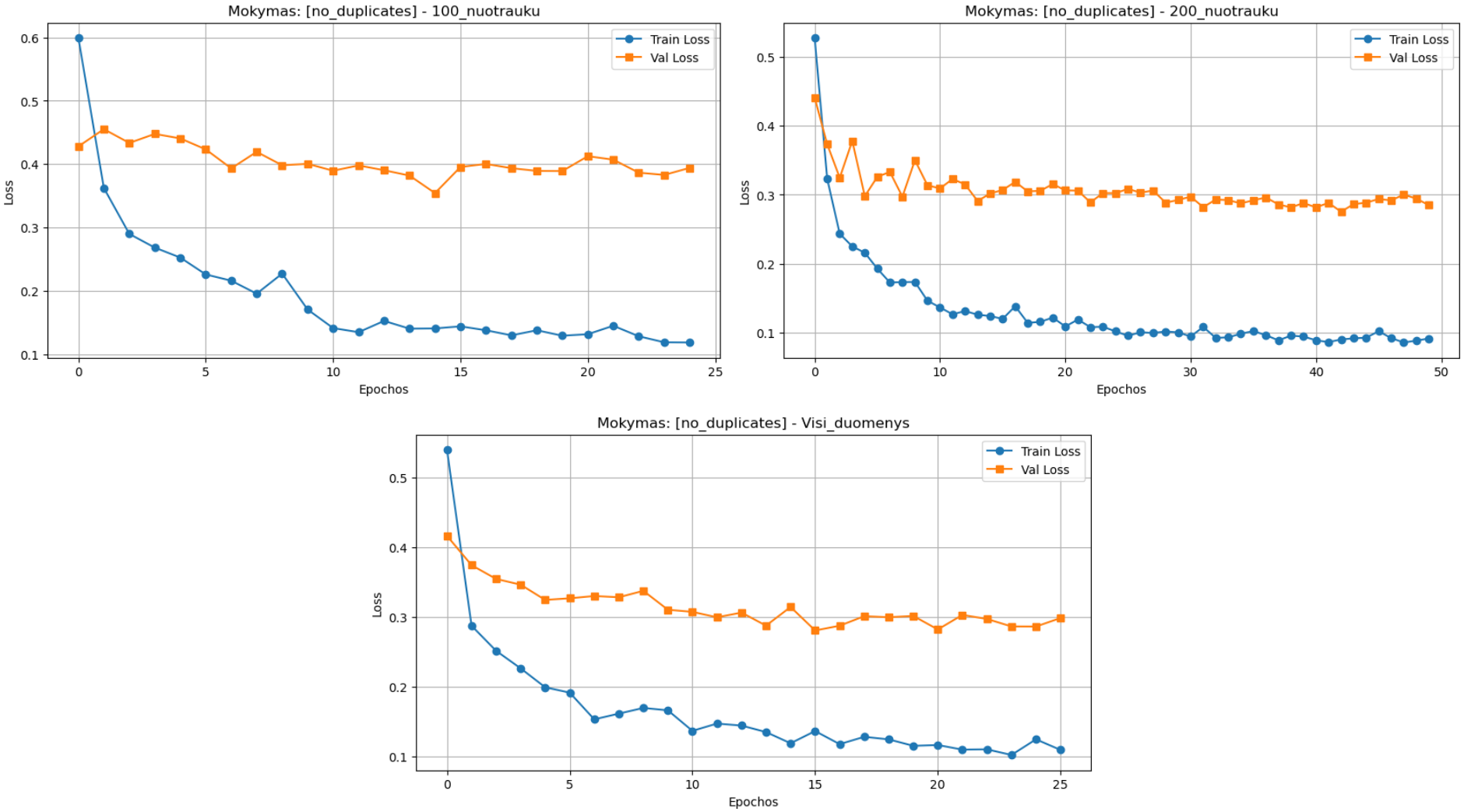
Šiame skyriuje aprašomos nematytų duomenų validacijos pirmo plano modelių mokymosi kreivės.

28 pav. lyginant mokymosi ir validacijos kreives su fono nematytų duomenų kreivėmis pastebimos panašios tendencijos. Naudojant mažiausią mokymo duomenų kiekį, validacijos rezultatai yra geresni už fono modelius, tačiau kreivės tarp modelių susivienodina naudojant visus galimus mokymo duomenis. Tai priklauso nuo jau minėtų validacijos aibės skirtumų ir dalinai dėl jau „pažįstamos“ logikos taikymo, kuria vadovaujantis skaidrūs objektai nėra ignoruojami, o aptinkami.

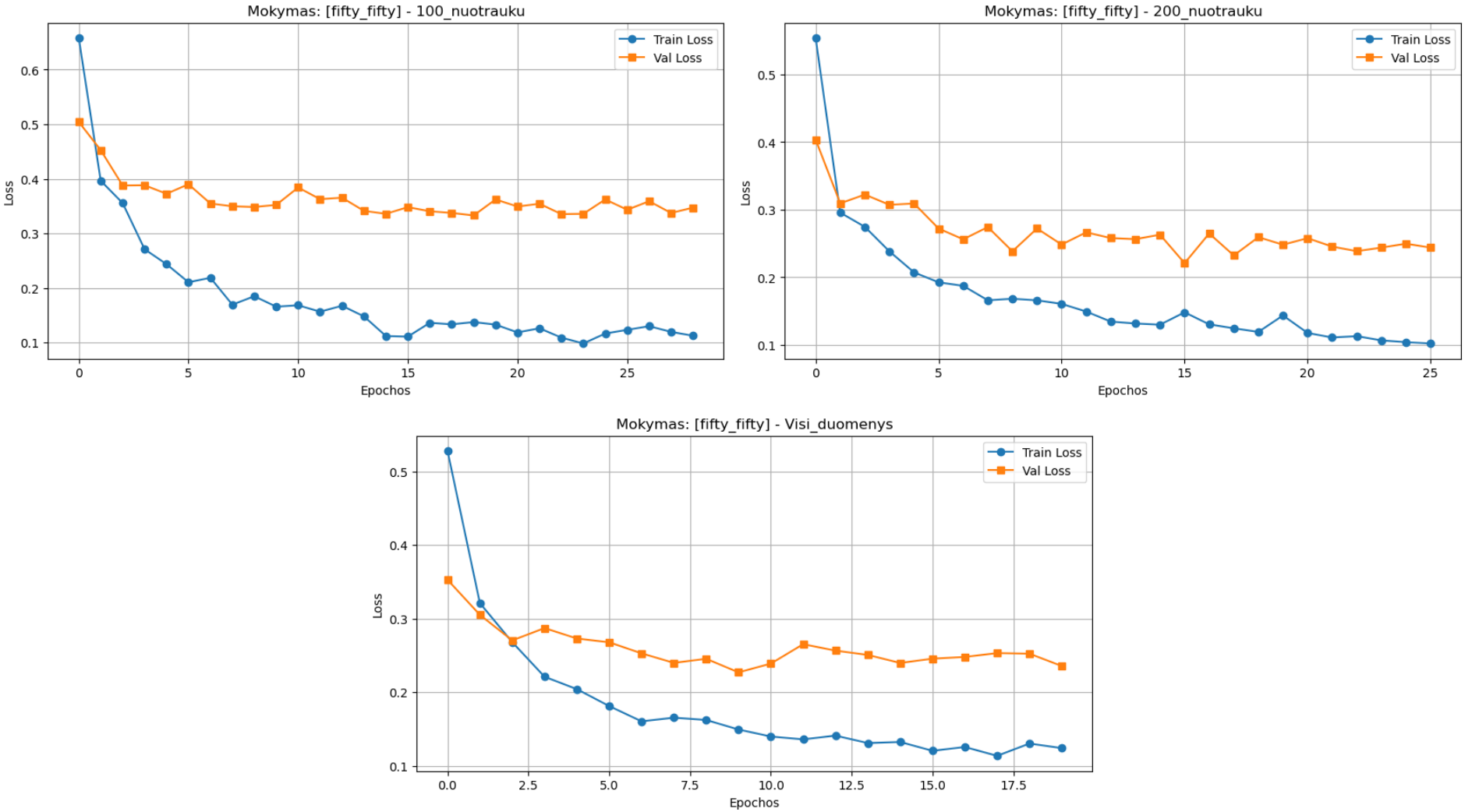
5) Hibridinė sistema (_50_50 modeliai)



27 pav. Matytų duomenų strategijos pirmo plano aptikimo modelių mokymosi kreivė



28 pav. Nematytų duomenų strategijos pirmo plano aptikimo modelių mokymosi kreivės



29 pav. Hibridinės strategijos pirmo plano aptikimo modelių mokymosi kreivės

29 pav. matomos kreivės yra panašios į fono modelio kreives. Validacijos kreivės ženkliai žemesnės už nematytos aibės, tačiau lyginant su geriausia etapo fono validacijos kreive (~0,15) pirmojo plano kreivės rezultatai prastesni (~0,22).

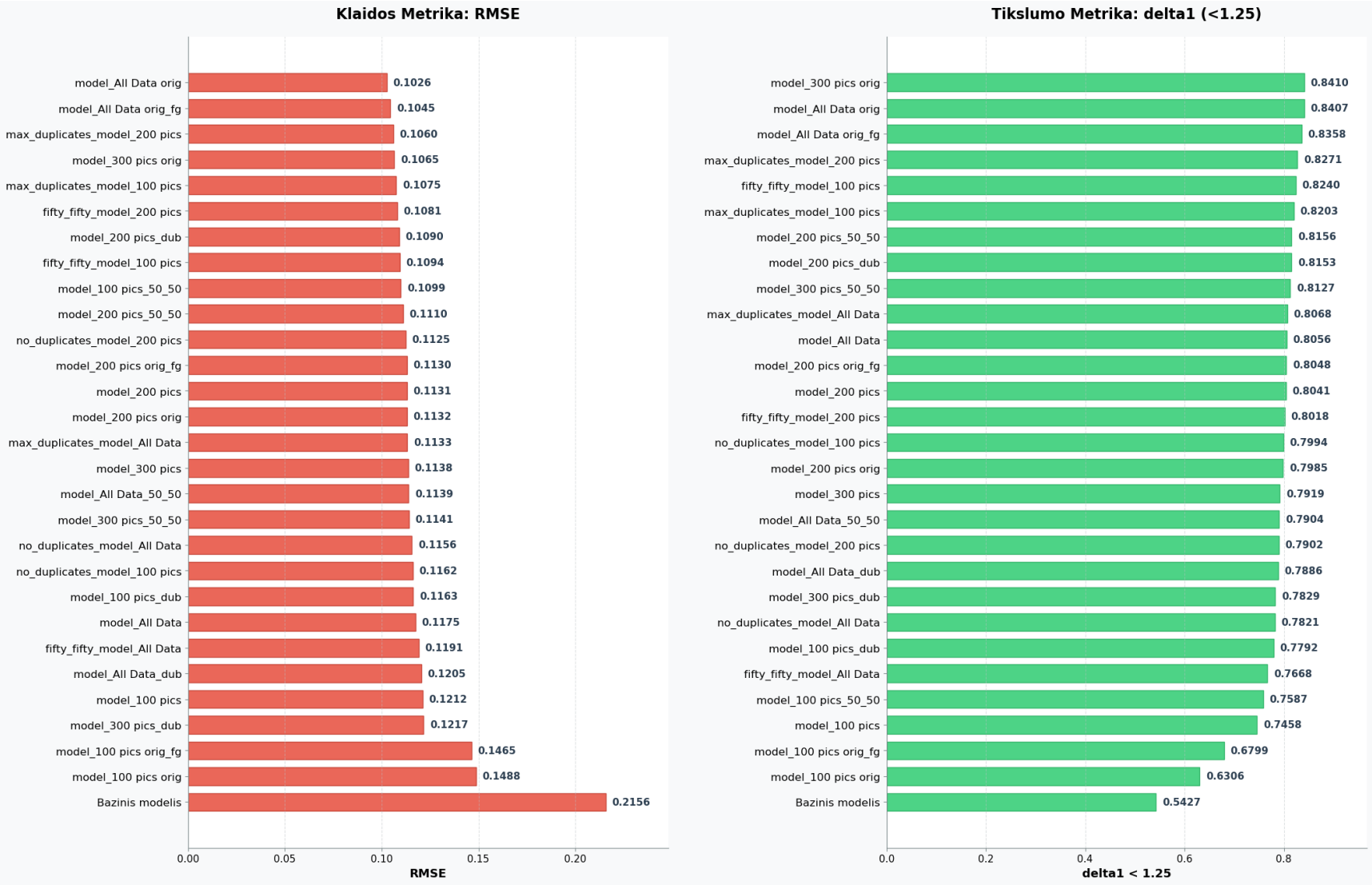
4.4. Rezultatų analizė

Dėl treniruotų modelių specifikacijų, standartinis testavimo procesas nepritaikomas. Pradinis vertinimo etapas įvertina modelio gebėjimą sudaryti gylio žemėlapius erdvėse, be skaidrių ar atspindžių objektų. Tai atliekama naudojantis DIODE duomenų rinkinio nenaudotomis scenomis (100 nuotraukų). Testavimo etapas palygina modelius tarpusavyje ir papildomai yra atlieka palyginamą su modeliu, kuris nebuvo treniruotas su DIODE duomenų rinkiniu. Tokio tipo palyginimas leidžia įvertinti duomenų rinkinio sukeliamą domeno poslinkį (angl. domain shift) ir patikrinti mokymo įtaka standartinių scenų prognozėms, kuomet naudojami modifikuoti duomenų rinkiniai.

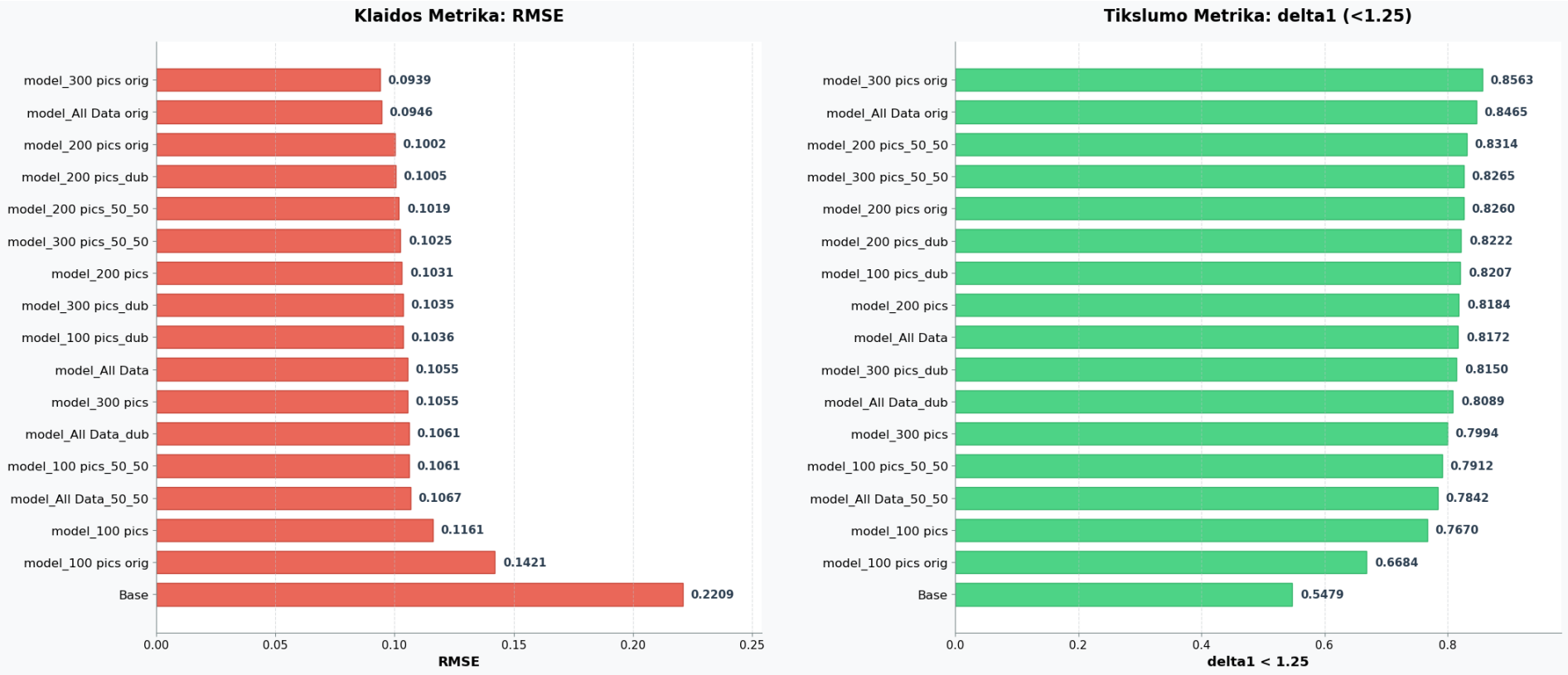
Fono aptikimo modeliai ir stiklinių paviršių aptikimo modeliai reikalauja papildomo testavimo atsižvelgiant į jų paskirtis. Fono modelis mokomas ignoruoti stiklinį objektą ir sudaryti gylio žemėlapį scenos, kuri vizualiai uždengta skaidraus objekto. Skaidraus objekto aptikimo modelis veikia priešingai ir turi sudaryti gylio žemėlapi tik skaidraus objekto o ne scenos, kuris dalinai matoma už jo. Tai reikalauja papildomos testavimo skilties, kuri atsižvelgtų ir įvertintu šiuos kritinius skirtumus.

Žiūrint į **30 pav.** vaizduojamos visų testuotų modelių RMSE ir delta1 < 1.25 vertės. Didžiausią įtaką vertėms turėjo modelio pritaikymas naujam domenui ir duomenų kiekis. Didžioji dalis modelių esančių viršutinėje paveikslo dalyje yra mokyti naudojantis didžiausiu kiekiu duomenų ir mokyti ant nemodifikuotų nuotraukų.

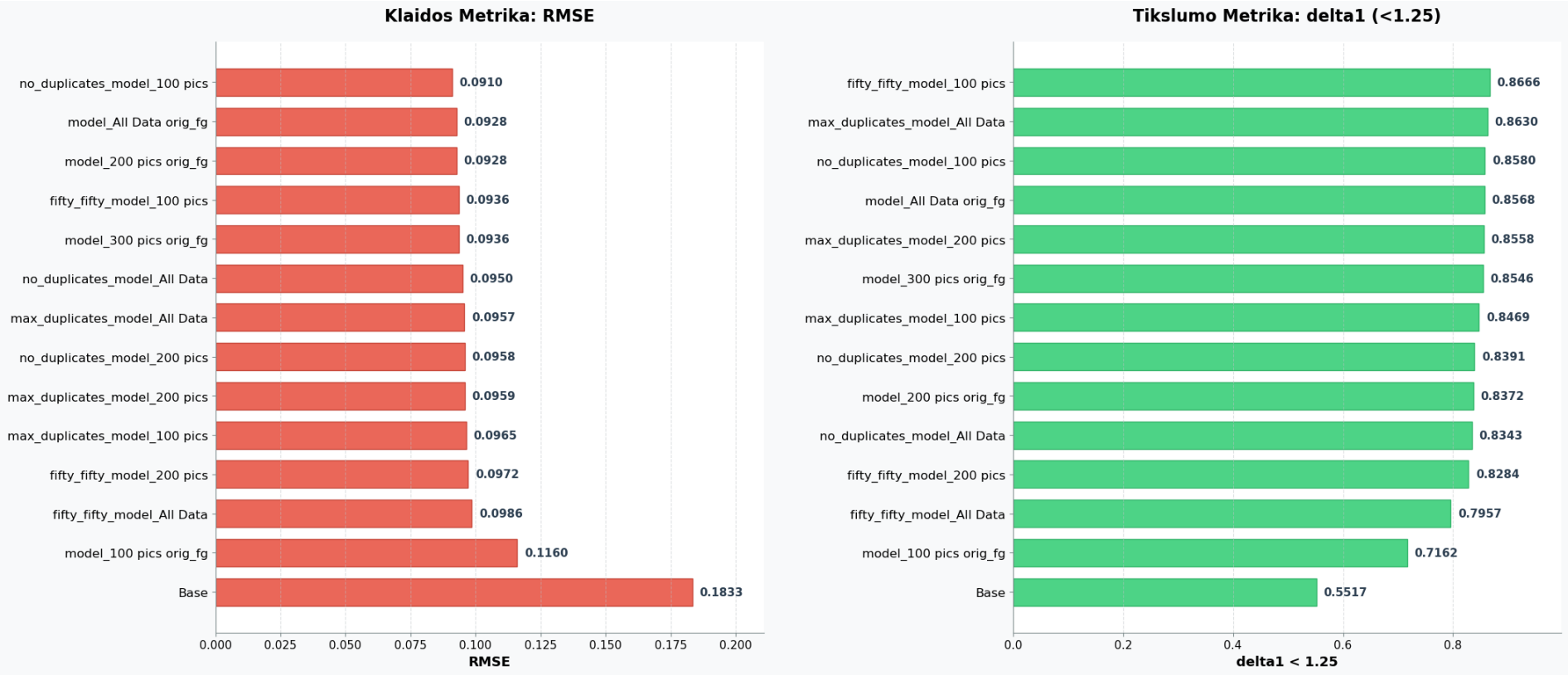
Siekiant lengviau vertinti modelius pagal specializaciją, **31 pav.** vaizduojami tik fono modelių ir bazinio modelio rezultatai. Paveikslėlyje geriausius įverčius pasiekę modeliai yra mokyti naudojant nemodifikuotas nuotraukas.



30 pav. Visų aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija



31 pav. Fono aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija



32 pav. Pirmo plano aptikimo modelių RMSE ir delta1 įverčių palyginimo vizualizacija

32 pav. vaizduojami pirmo plano aptikimo modelių rezultatai. Geriausius įverčius pasiekę modeliai yra „*no_duplicates_model_100*" ir „*fifty_fifty_model_100*". RMSE rezultatai vis tiek palankesni nemodifikuotiems vaizdams, tačiau delta vertės indikuoja geresnius pasiekimus modelių, kurie apmokti su modifikuotais vaizdais.

4.4.1. Gylio žemėlapių kokybės įvertinimas standartinėse scenose

Modeliai įvertinami naudojant 100 DIODE duomenų rinkinio nuotraukas, kurios nebuvo naudotos apmokymo ar validacijos procesuose. Remiantis prielaida, kad skaidrūs kūnai scenose neegzistuoja, galima vertinti pirmojo plano ir fono modelius kartu su baziniais modeliais.

Pateikta **Error! Not a valid bookmark self-reference.** atvaizduoja modelių klaidų vertinimo rezultatus. Visos pateiktos metrikos (RMSE, AbsRel, SqRel, Log10) nurodo paklaidą. Lentelė yra surikiuota pagal RMSE metriką nuo geriausio iki prasčiausio modelio.

Analizuojant lentelę, labiausiai matoma išskirtis yra pačioje apačioje. „*Bazinis modelis*" yra ženkliai prastesnis visose kategorijose (RMSE: 0,2156). Visi modelių mokymo etapai pagerino modelių veikimą. Geriausias modelis („*model_Visi_duomenys_originalai*") klaidos rodiklį RMSE sumažino daugiau nei perpus (nuo 0,2156 iki 0,1026). Tai pagrindžia pasirinkto sprendimo adaptacijos tikslingumą.

Mažiausiais RMSE įverčiais pasižymi modeliai , treniruoti su visais originaliais duomenimis:

1. „*model_Visi_duomenys_originalai*" (RMSE: 0,1026)
2. „*model_Visi_duomenys_originalai_fg*" (RMSE: 0,1045)

Modeliai mokyti naudojant didelį kiekį originalių nemodifikuotų duomenų pasiekė geresnius rezultatus už visas kitas strategijas. Tai patvirtina klasikinę mašininio mokymosi taisyklę: aukštos kokybės, nemodifikuoti duomenys duoda stabiliausius rezultatus baziniam gylio nustatymui.

6 lentelėje matosi aiški koreliacija tarp treniravimui naudotų nuotraukų skaičiaus ir modelio tikslumo. Modeliai, treniruoti tik su 100 nuotraukų, yra sąrašo apačioje. Naudojantis 200-300 nuotraukų rezultatai stabilizuojasi ir gerėja. 100 nuotraukų neužtenka norint pasiekti norimą/tenkinamą rezultatą, tačiau perėjimas nuo 300 prie "Visų duomenų" nesukuria tokio paties ženklaus skirtumo.

Nors originalūs duomenys suteikė geriausius rezultatus, verta atkreipti dėmesį į „*max_duplicates_model_200_nuotrauku*" modelį. Jo RMSE įvertis yra trečias geriausias rezultatas (RMSE: 0,1060) ir yra geresnis nei strategija, naudojanti didesnį kiekis originalių duomenų. Tai parodo, kad specifinės duomenų augmentacijos ir dubliavimo technikos gali kompensuoti mažesnį duomenų kiekį.

Kiti stulpeliai (AbsRel, SqRel, Log10) beveik idealiai atkartoja RMSE tendencijas. Pavyzdžiui, geriausias RMSE modelis taip pat turi ir mažiausią absoliučią santykinę klaidą (AbsRel: 0,132). Modelių patobulėjimas nėra tik vienos metrikos atžvilgiu. Modeliai, kurie geriausiai išvengia didelių klaidų, lygiai taip pat gerai veikia bendro procentinio tikslumo atžvilgiu.

6 lentelė. Klaidos metrikų įverčiai

Modelis	RMSE	AbsRel	SqRel	Log10
model_Visi_duomenys_originalai	0,1026	0,132	0,0234	0,0637
model_Visi_duomenys_originalai_fg	0,1045	0,1367	0,0252	0,0649
max_duplicates_model_200_nuotrauku	0,106	0,1369	0,0252	0,0614
model_300_nuotrauku_originalai	0,1065	0,1311	0,0254	0,0601
max_duplicates_model_100_nuotrauku	0,1075	0,1432	0,027	0,0675
fifty_fifty_model_200_nuotrauku	0,1081	0,1491	0,0245	0,074
model_200_nuotrauku_dub	0,109	0,1445	0,0293	0,071
fifty_fifty_model_100_nuotrauku	0,1094	0,1376	0,0267	0,0618
model_100_nuotrauku_50_50	0,1099	0,1616	0,0271	0,0806
model_200_nuotrauku_50_50	0,111	0,1407	0,0268	0,0647
no_duplicates_model_200_nuotrauku	0,1125	0,1556	0,029	0,0735
model_200_nuotrauku_originalai_fg	0,113	0,1533	0,0282	0,0772
model_200_nuotrauku	0,1131	0,1488	0,0304	0,0707
model_200_nuotrauku_originalai	0,1132	0,1564	0,0318	0,0763
max_duplicates_model_Visi_duomenys	0,1133	0,149	0,0304	0,0683
model_300_nuotrauku	0,1138	0,1577	0,0314	0,0819
model_Visi_duomenys_50_50	0,1139	0,1542	0,029	0,0785
model_300_nuotrauku_50_50	0,1141	0,1469	0,0302	0,0682
no_duplicates_model_Visi_duomenys	0,1156	0,1616	0,0334	0,0811
no_duplicates_model_100_nuotrauku	0,1162	0,1522	0,0321	0,0688
model_100_nuotrauku_dub	0,1163	0,1551	0,0289	0,0761
model_Visi_duomenys	0,1175	0,147	0,0333	0,0666
fifty_fifty_model_Visi_duomenys	0,1191	0,1691	0,0328	0,0884
model_Visi_duomenys_dub	0,1205	0,155	0,0331	0,0705
model_100_nuotrauku	0,1212	0,1713	0,0312	0,0859
model_300_nuotrauku_dub	0,1217	0,1591	0,0347	0,0757
model_100_nuotrauku_originalai_fg	0,1465	0,207	0,0509	0,096
model_100_nuotrauku_originalai	0,1488	0,2262	0,0486	0,1065
Bazinis modelis	0,2156	0,285	0,1389	0,1389

7 lentelė. Tikslumo metrikų įverčiai

Modelis	$\delta_1 < 1.25$	$\delta_2 < 1.25^2$	$\delta_3 < 1.25^3$
model_300_nuotrauku_originalai	0,841	0,9377	0,9697
model_Visi_duomenys_originalai	0,8407	0,934	0,963
model_Visi_duomenys_originalai_fg	0,8358	0,9367	0,9681
max_duplicates_model_200_nuotrauku	0,8271	0,9353	0,9709
fifty_fifty_model_100_nuotrauku	0,824	0,9377	0,9732
max_duplicates_model_100_nuotrauku	0,8203	0,9272	0,9623
model_200_nuotrauku_50_50	0,8156	0,9292	0,9664
model_200_nuotrauku_dub	0,8153	0,9293	0,9562
model_300_nuotrauku_50_50	0,8127	0,9205	0,9635
max_duplicates_model_Visi_duomenys	0,8068	0,9326	0,9633
model_Visi_duomenys	0,8056	0,9339	0,9661
model_200_nuotrauku_originalai_fg	0,8048	0,9145	0,9567
model_200_nuotrauku	0,8041	0,924	0,9596
fifty_fifty_model_200_nuotrauku	0,8018	0,9248	0,9587
no_duplicates_model_100_nuotrauku	0,7994	0,93	0,963
model_200_nuotrauku_originalai	0,7985	0,9123	0,9549
model_300_nuotrauku	0,7919	0,9117	0,9458
model_Visi_duomenys_50_50	0,7904	0,915	0,9526
no_duplicates_model_200_nuotrauku	0,7902	0,9264	0,9589
model_Visi_duomenys_dub	0,7886	0,926	0,9645
model_300_nuotrauku_dub	0,7829	0,9143	0,9572
no_duplicates_model_Visi_duomenys	0,7821	0,9112	0,9526
model_100_nuotrauku_dub	0,7792	0,9217	0,9584
fifty_fifty_model_Visi_duomenys	0,7668	0,8938	0,9374
model_100_nuotrauku_50_50	0,7587	0,9195	0,957
model_100_nuotrauku	0,7458	0,9047	0,9522
model_100_nuotrauku_originalai_fg	0,6799	0,8635	0,9327
model_100_nuotrauku_originalai	0,6306	0,8412	0,92
Bazinis modelis	0,5427	0,7738	0,8791

7 lentelėje matomos modelių tikslumo metrikos ($\delta_1, \delta_2, \delta_3$). Priešingai nei prieš tai buvusioje klaidų lentelėje (RMSE), čia didesnė reikšmė reiškia geresnį rezultatą. Lentelė yra surikiuota pagal griežčiausią metriką $\delta_1 < 1.25$. Kaip ir klaidų lentelėje, „*Bazinis modelis*“ yra pačioje lentelės apačioje(bazinio modelio δ_1: 0,5427). Tai reiškia, kad tik apie 54 % nuotraukos atstumų prognozuota pakankamai tiksliai. Geriausias treniruotas modelis šį skaičių aplenkia iki ~0,84 (84 %). Tai yra 30 procentų geresnis rezultatas.

Viršuje lentelės vėl dominuoja originalūs duomenys, tačiau matomas labai svarbus skirtumas:

1. „*model_300_nuotrauku_originalai*” : 0,8410
2. „*model_Visi_duomenys_originalai*” : 0,8407

Modelis, treniruotas su 300 nuotraukų, aplenkė modelį treniruotą su visais duomenimis.

Toliau analizuojant lentelę matomas svarbus rezultatas lyginant skirtingas maišymo strategijas. Modelio „*model_100_nuotrauku_originalai*“ rezultatas antras nuo galo (0,6306). Tačiau pritaikius maišymo strategiją tam pačiam 100 nuotraukų kiekiui, rezultatai išauga iki pat lentelės viršaus: „*fifty_fifty_model_100_nuotrauku*“ (0,8240) penktoje vietoje. Žvelgiant į δ_3 stulpelį, matoma, kad visi treniruoti modeliai pasiekė virš 92 % tikslumą.

Nors didžiausias originalių duomenų kiekis užtikrina mažiausią vidutinę klaidą (RMSE), modelio tikslumo augimas sustoja ties 300 nuotraukų. Svarbu paminėti, jog rezultatai taip pat parodė, kad siūlomos alternatyvios hibridinės mokymo strategijos yra veiksmingos tais atvejais, kai originalių duomenų kiekis yra labai mažas.

4.4.2. Specializuotų modelių įvertinimas originaliame domene

Siekiant patikrinti modelių sąveiką su skaidriais ir atspindžiais paviršiais atlikti papildomi bandymai su nematytomis DIODE duomenų rinkinio scenomis. Papildant sceną naujais skaidriais objektais, kurie uždengia didelę dalį scenos, įvertinamas pirmojo plano aptikimo modelio gebėjimas matyti skaidrius kūnus ir fono modelio antro plano geometrijos prognozavimas. Lyginant gautus gylio žemėlapius su turimais nemodifikuotų scenų gylio žemėlapiais, galima įvertinti specializuotų modelių efektyvumą. Fono aptikimo modelio sukurtas scenos prognozuojamas gylio žemėlapis turėtu atitikti originalą su minimalia paklaida. Naudojant modelį treniruotą tik ant originalių duomenų tikimasi standartinės paklaidos kuri sukeliama skaidrių kūnų. Dalis skaidraus kūno bus vertinama kaip esanti arti kameros, o kita dalis bus kiaurai permatoma ir bus nuspėjama geometrija už stiklinio kūno. Tačiau pritaikius modelį, specializuotą į skaidrių kūnų aptikimą, tikimasi gauti visą skaidraus kūno gylio žemėlapį kaip plokštumą, ir taip išvengti standartinės skaidrių kūnų dviprasmybės.

Modeliams įvertinti buvo pasirinktos 10 prieš tai nenaudotų DIODE duomenų rinkinio scenų. Scenos buvo modifikuotos naudojantis Google Gemini Pro Image. Gemini modeliui pateikta instrukcija:

„Add a glass plane on the left side of the image. It should be barely visible and not change the scenes original lay out at all. Glass should start at the middle of the picture.“.

Instrukcija leidžia uždengti didelę dalį scenos aiškiai padalinti sceną į dvi dalis. Kairėje pusėje sugeneruota skaidri siena sudaro pusę viso vaizdo ir modeliams, kurie nėra specializuoti ignoruoti skaidrius kūnus tai patampa neįmanoma užduotis.

Testavimui parinkti modeliai:

1. Pirmo plano modelis: „*max_duplicates_model_200_nuotrauku*“
2. Fono modelis: „*model_200_nuotrauku_dub*“
3. Originalių vaizdų modelis: „*model_Visi_duomenys_originalai*“
4. DepthAnythingV2 netreniruotas bazinis modelis.

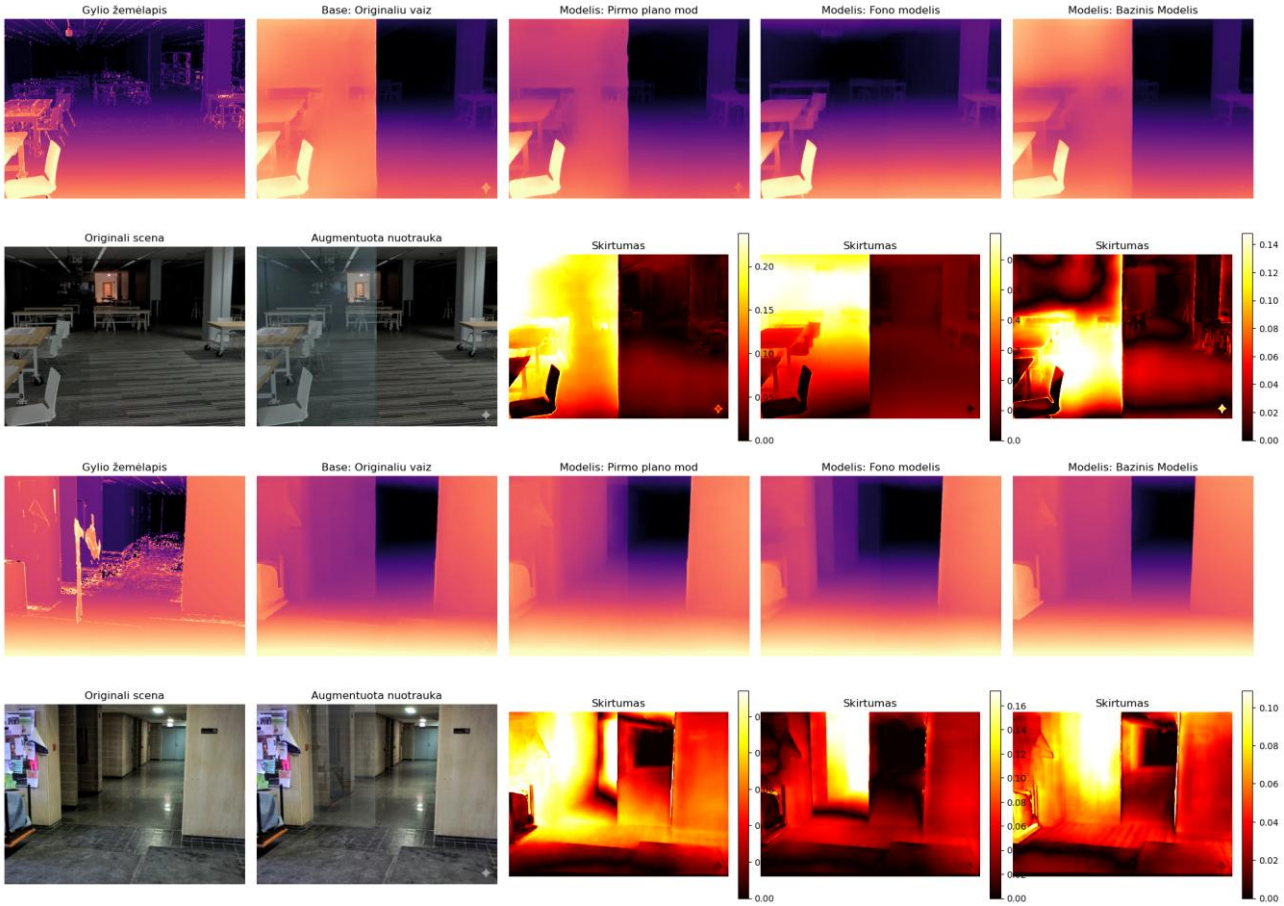
8 lentelėje matomi 10 nuotraukų rezultatai. Bazinis modelis, kuris nebuvo mokytas prie DIODE duomenų rinkinio, pasirodė prasčiausiai. Fono modelis pasiekė geriausius rezultatus, dėl savo gebėjimo matyti kiaurai skaidrius objektus ir sudaryti gylio žemėlapius už jų. Pirmo plano modelis pasirodė prasčiausiai iš trijų mokytų modelių.

8 lentelė. Dešimt modifikuotų nematytų scenų metrikų įverčiai

Metrikos	Bazinis Modelis	Originalių vaizdų modelis	Pirmo plano modelis	Fono modelis
RMSE	0.1368	0.0894	0.0905	0.0785
AbsRel	0.4113	0.2520	0.2513	0.1966
SqRel	0.0781	0.0326	0.0423	0.0198
Log10	0.2125	0.1229	0.1035	0.0998
$\delta_1 < 1.25$	0.4692	0.7113	0.7024	0.7314
$\delta_2 < 1.25^2$	0.6748	0.8350	0.8288	0.8528
$\delta_3 < 1.25^3$	0.7875	0.8787	0.9089	0.9264

Pilnai įvertinti modelius tik pagal metrikų reikšmes negalima. Mokant modelius atlikti sudėtingą uždavinį, kaip skaidrių kūnų ignoravimas ar vaizdavimas, reikia papildomos vizualinės patikros. Tai leidžia įvertinti modelius pagal jų gebėjimą susidoroti su skaidriais kūnais pagal paskirtį.

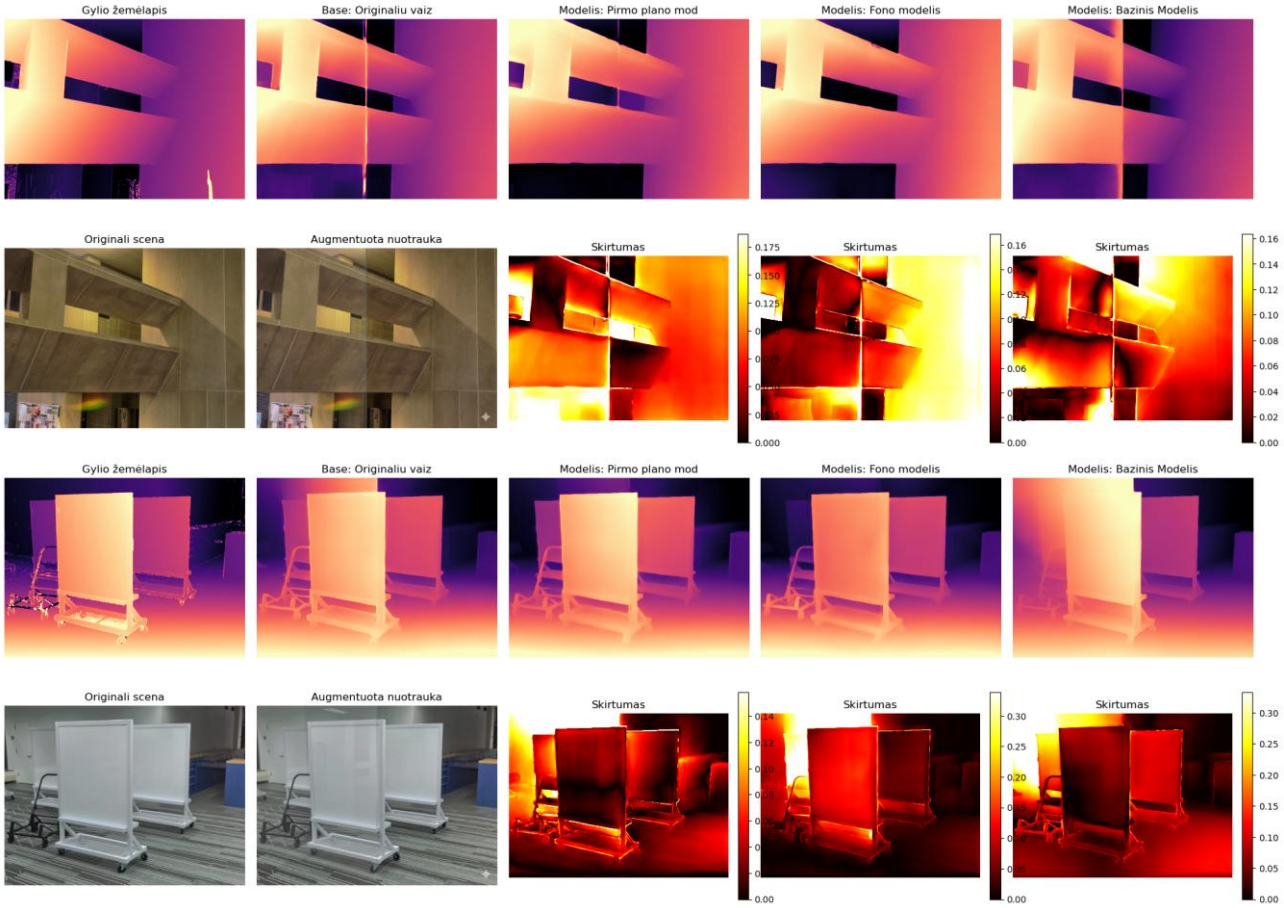
33 pav. vaizduojamos dvi skirtingos scenos ir modelių suformuoti gylio žemėlapiai. Pirmoje ir trečioje eilutėse matomas originalus gylio žemėlapis priklausantis scenai ir visų testuotų modelių prognozės. Matomas ženklus modelių skirtumas kairėje scenos dalyje, kurioje buvo sintetiškai įklijuotas skaidrus kūnas. Toks pavyzdys gerai iliustruoja prieš tai minėtą nelambertinių kūnų sukeliamą dviprasmybę. Trys modeliai prognozuoja ir skaidraus kūno plokštumą ir objektų už jos atstumą. Antroje ir ketvirtoje eilutėse matomas originalus ir modifikuotas scenos vaizdas, bei skirtumo žemėlapiai. Skirtumo žemėlapiai sudaromi skaičiuojant skirtumą tarp „*Base*“ sudaryto gylio žemėlapio ir „*Modelis*“ sudaryto gylio žemėlapio. Skirtumų žemėlapiai leidžia vizualiai matyti didžiausius neatitikimus tarp skirtingų modelių prognozių.



33 pav. Modifikuotų scenų pavyzdžiai, kuriuose labiausiai pastebimas sudarytų gylio žemėlapių skirtumas

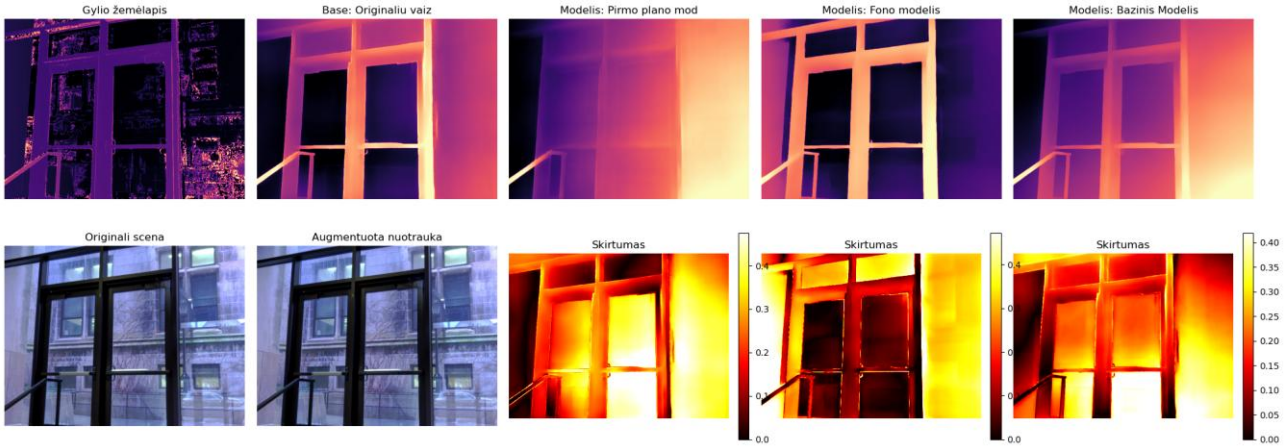
34 pav. vaizduojamos dvi skirtingos scenos ir modelių suformuoti gylio žemėlapiai. Pirmoje ir trečioje eilutėse matomas originalus gylio žemėlapis priklausantis scenai ir visų testuotų modelių prognozės. Visi modeliai neaptiko skaidrios plokštumos.

Net analizuojant bazinio modelio sudarytą gylio žemėlapį, susiduriama su minimaliu kiekiu skaidraus kūno sukeliamais artefaktais. Antroje ir ketvirtoje eilutėse matomas originalus ir modifikuotas scenos vaizdas, bei skirtumo žemėlapiai. Skirtumo žemėlapiai nepabrėžia visos stiklu dengtos scenos dalies. Skirtumai pagrinde matomi stiklo plokštumos pradžioje, kuri išskiria sceną į dvi dalis, o ne visame plokštumos plote.



34 pav. Modifikuotų scenų pavyzdžiai, kuriuose visi modeliai prognozavo fono, o ne stiklo plokštumos, gylio žemėlapius

35 pav. vaizduojama nemodifikuota scena iš DIODE duomenų rinkinio. Pavyzdys iliustruoja pagrindinę problemą, su kuria susiduriama scenai vaizduojant skaidrius kūnus. Visi modeliai suformavo radikaliai besiskiriančius gylio žemėlapius. Žvelgiant į gylio žemėlapį, kuris buvo suformuotas kuriant DIODE duomenų rinkinį, matomi klaidingi duomenys. Dėl pusiau atspindinčio stiklo paviršiaus, kai kurie taškai gražinami arčiau negu kiti. Tai sukuria neteisingus mokymo duomenis ir išskirtis, kurios trikdo modeliams mokytis.

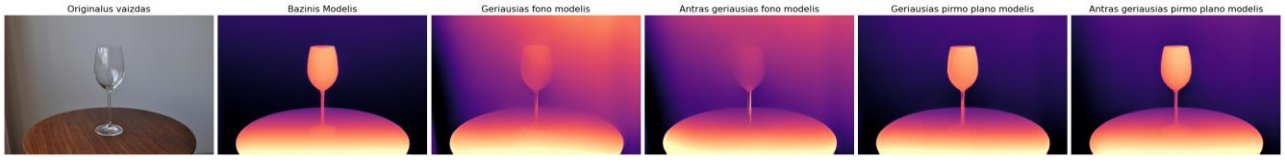


35 pav. DIODE duomenų rinkinio scena su dideliu kiekiu nelambertinių paviršių. Skirtingų modelių gylio žemėlapiai radikaliai skiriasi vienas nuo kito ir nuo tikrojo jutiklio gauto atstumo

Tikslesniems ir platesniems testavimo etapams reikia specializuotų duomenų rinkinių. Didesnį nuotraukų skaičius privestų prie tikslesnių įverčių ir detalesnės modelių veikimo rezultatų analizės.

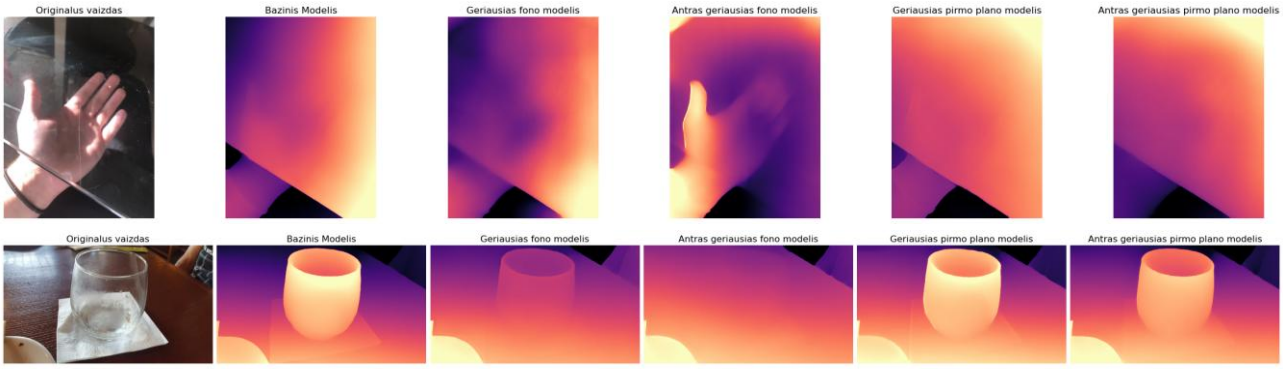
4.4.3. Specializuotų modelių įvertinimas realaus pasaulio aplinkoje

Siekiant įvertinti modelius realaus pasaulio aplikacijose naudojantis skirtingais fotoaparatais ir aplinkomis, pasitelktos atsitiktinai internete išrinktos testavimo nuotraukos. Nuotraukos skiriasi raiška, fotoaparatų parametrais, apšvietimo sąlygomis ir kai kurios netgi turi vandens žymias.



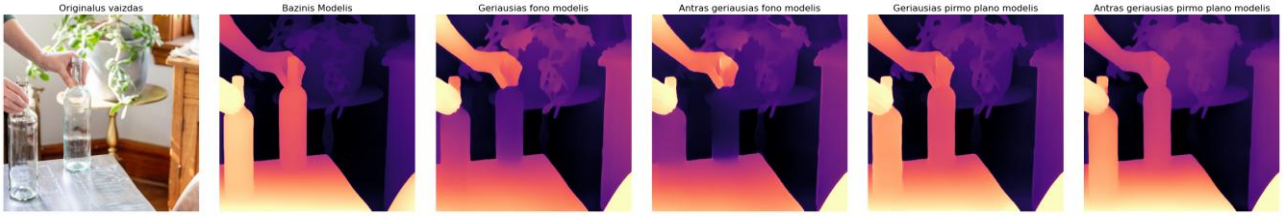
36 pav. Realaus pasaulio scena su taure ir stalu (Nuotaukos šaltinis: [49])

36 pav. vaizduojama nuotrauka su skaidriu objektu. Bazinis modelis yra naudojamas kaip atskaitos taškas specializuotiems modeliams vertinti. Pakeitus pritaikymo domeną, naudojamas nemokytas DepthAnythingV2 bazinis modelis. Pateiktame pavyzdyje matomas neteisingas fono modelių rezultatas. Vietoj skaidraus modelio ignoravimo, objekto siluetas išlieka, bet jis suvienodinamas su fono prognoze. Fonas prognozuojamas neteisingai ir prarandama aiški scenos struktūra, kuri yra matoma pirmo plano ir bazinių modelių prognozėse. Pirmo plano modelių žemėlapiai atrodo vizualiai identiški į bazinio modeliu ir neturi radikalių skirtumų tarp jų.



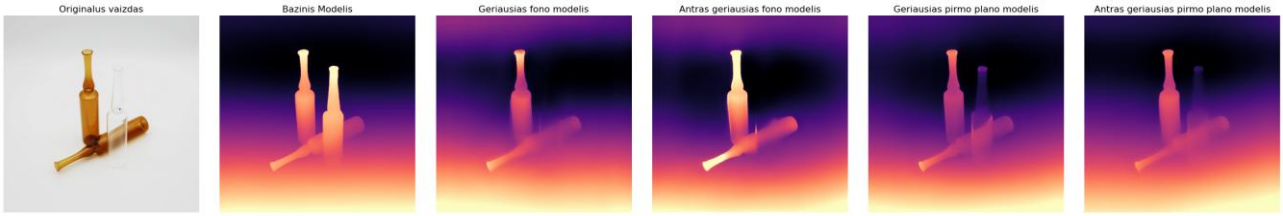
37 pav. Realaus pasaulio scenos su stiklo plokšte ir stikline (Nuotaukų šaltiniai: [50, 51])

37 pav. vaizduojamos scenos kuriuose fono ir pirmo plano modeliai pasirodė taip, kaip tikėtasi, pavyzdžiuose išryškėja poreikis vaizduoti ne vien geriausiai įvertintus modelius, bet ir kitus modelius, šiuo atveju antrus pagal įverčius. Geriausiame fono modelio sudarytuose gylio žemėlapiuose, delnas nėra matomas, o taurė yra. Tačiau žvelgiant į antro geriausio fono modelio gylio žemėlapius, matomas delnas ir individualūs pirštai, taurė išnyksta pilnai ir lieka tik stalo plokštuma. Pirmojo plano modeliai prognozuoja atstumus teisingai ir išlaiko stiklo plokštumos geometriją geriau už bazinį modelį.



38 pav. Realaus pasaulio scenarijus su dviem stikliniais buteliais (Nuotraukos šaltinis: [52])

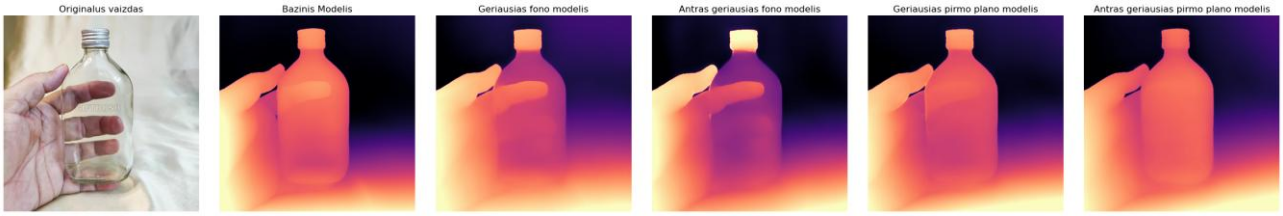
38 pav. vaizduojamoje scenoje yra du skaidrūs objektai. Antram geriausiam fono modeliui geriau pavyksta sudaryti gylio žemėlapį, tačiau sukuriamos kelios kritinės klaidos. Butelis esantis arčiau kameros nėra pilnai ištrinamas, o paliekamas kaip stalo paviršiaus tęsinys. Antra klaida matoma žvelgiant virš rankų, kurios laiko butelį. Butelių kakleliams ir rankoms yra priskiriami tokie patys atstumo gyliai. Fono modeliai sudaro detalesnį augalo gylio žemėlapį, o pirmojo plano modeliai prognozuoja panašiai, kaip bazinis modelis.



39 pav. Realaus pasaulio scenarijus su rudomis ir skaidriomis ampulėmis (Nuotraukos šaltinis: [53])

39 pav. yra matoma sudėtingesnė scena, kuri turi spalvoto ir nespalvoto stiklo objektų. Visi objektai matomi scenoje yra skaidrūs, tačiau ne vienodo skaidrumo. Fono modeliai sukuria gylio žemėlapius be skaidriausio objekto, tačiau rudo stiklo ampules aptinka ir prognozuoja kaip ne skaidrius objektus. Pirmojo plano modeliai visiškai suklysta ir sudaro prastesnius gylio žemėlapius lyginant su baziniu

modeliu. Modeliai beveik ištrina skaidrų objektą, tačiau ryškiai matomas jo siluetas ir nevienodo atstumo gyliai.



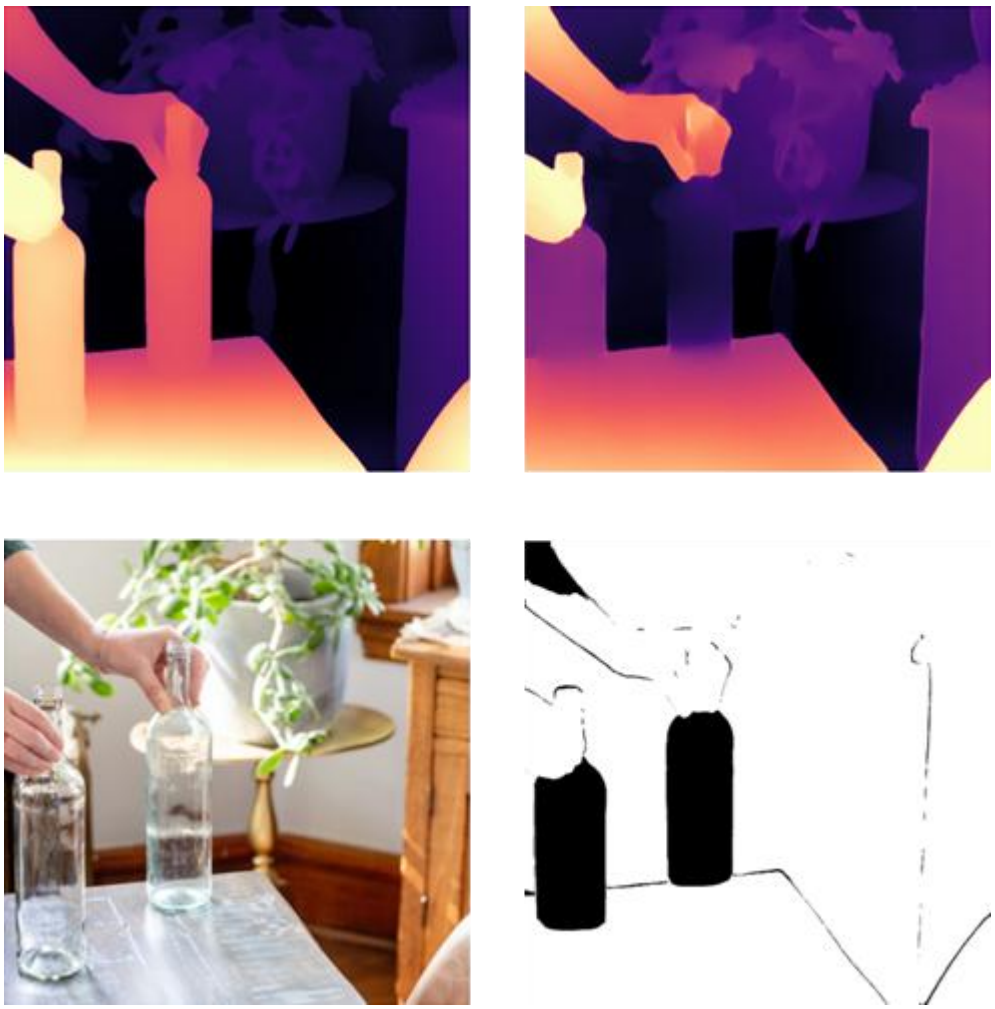
40 pav. Realaus pasaulio scena su skaidriu buteliu ir delnu (Nuotraukos šaltinis: [54])

40 pav. vaizduojamoje scenoje yra matomas skaidrus butelis ir jį laikantis delnas. Fono modeliai išskiria pirštus geriau nei bazinis modelis. Antras geriausias pirmojo plano modelis neišskiria pirštų ir prognozuoja vientisą stiklinio butelio plokštumą, tačiau praranda delno formos tikslumą.

Vizualiai įvertinus skirtingus modelius ne domeno aplinkoje pabrėžiamos skirtingų modelių veikimo klaidos ir teisingai prognozuotos scenos. Visi modeliai suklydo bent vienoje scenoje. Priklausomai nuo scenos aplinkybių, modelių rezultatai skiriasi ir antri geriausi modeliai prognozuoja taisyklingesnius gylio žemėlapius.

4.5. Modelių pritaikymas

Išanalizavus modelių veikimo ypatumus, matomos potencialios jų pritaikymo sritys. Naudojant abu modelius scenoje su skaidriais kūnais, sukurti gylio žemėlapiai ryškiausiai skiriasi vietose, kuriose yra skaidrūs paviršiai arba paviršiai su atspindžiais. Pasirinkus skirtumo ribą ir apdorojus skirtumų žemėlapį, galima sugeneruoti binarinę segmentavimo kaukę (žr. **41 pav.**), kuri pažymi skaidrius kūnus scenoje. Tokio tipo kaukė galėtu būti naudojama kaip įtartinų arba apgaulingų scenos vietų žymeklis, leidžiantis sistemai išvengti neteisingų arba dviprasmių atstumo rodmenų.



41 pav. Palyginus sugeneruotus gylio žemėlapius naudojantis pirmo ir antro plano modelius, gaunama binarinė segmentacijos kaukė padengianti skaidrius butelius (Nuotraukos šaltinis: [52])

Modelių gebėjimas prognozuoti fono geometriją gali būti pritaikytas ne vien skaidriems kūnams. Vizualią informaciją iš dalies uždengiančios kliūtys egzistuoja skirtingais formatais. Modelių veikimo stilistika galėtų būti taikoma lietaus, rūko, pūgos ar net smėlio audros oro sąlygoms įveikti. Nors darbe atlikti eksperimentai buvo riboti duomenų kiekiu ir resursais, tačiau parodė potencialias modelių pritaikymo sritis ir galimybes.

Išvados

1. Mokslinės literatūros analizė atskleidė, kad monokuliarinio gylio nustatymo (MDE) modeliai pasižymi didelėmis paklaidomis analizuodami nelambertinius paviršius dėl specifinių vaizdinių žymių (skaidrumo, atspindžių) trūkumo mokymo duomenyse. Nustatyta, kad generatyvinio dirbtinio intelekto technologijos suteikia galimybę kurti aukštos kokybės sintetinius duomenis, kurie gali kompensuoti tradicinių jutiklių, tokių kaip LiDAR, fizinius ribotumus fiksuojant skaidrius objektus.
2. Pasitelkiant multimodalinius bei difuzinius modelius sukurti specializuoti sintetinių duomenų rinkiniai, kuriuose naudojant generatyvinius modelius buvo simuliuojamos scenos su skaidriais objektais. Pasiūlyta metodika generuoja ne tik vizualiai korektiškus vaizdus, bet ir jiems priskirtą fono bei objektų geometriją atspindinčius gylio žemėlapius, būtinus neuroninių tinklų apmokymui.
3. Apmokius gylio nustatymo modelius su „DIODE“ ir sintetiniais duomenų rinkiniais, nustatyta, kad sugeneruotų duomenų kokybė daro įtaką modelių tikslumui. Pirmojo plano (skaidrių objektų aptikimo) modelio rezultatai parodė, kad per didelis netinkamų duomenų kiekis mokymo imtyje lemia gylio žemėlapių detalumo praradimą ir sumažina modelio gebėjimą atpažinti šiuos objektus. Tuo tarpu fono gylio nustatymo (skaidrių kūnų ignoravimo) modelis, apmokytas kokybiškesniais duomenimis, veikia patikimiau – jis tiksliau nustato scenos geometriją vietose, kurias iš dalies uždengia skaidrūs objektai.
4. Apmokytų modelių vertinimas buvo atliktas dviem etapais: apskaičiuojant tikslumo metrikas originalaus duomenų rinkinio domene ir atliekant vizualinę analizę su naujomis (ne domeno) scenomis. Kiekybinė analizė parodė, kad fono gylio nustatymo modelis pasiekė gerus rezultatus (RMSE paklaida 0,1005; δ_1 tikslumas 0,8222) – jis reikšmingai aplenkė bazinį netaikytą modelį (RMSE 0,2209), nors nusileido modeliui, mokytam tik su nemodifikuotais duomenimis (RMSE 0,0939). Pastebėta tai, kad vizualinė analizė atskleidė svarbų neatitikimą: praktikoje sklandžiausiai stiklo objektus ignoruojantys fono modeliai pagal metrikas užėmė tik antrąją vietą, tačiau pademonstravo didžiausią potencialą kurti vientisus gylio žemėlapius. Tuo tarpu pirmojo plano modelis pagal metrikas pasiekė pačius geriausius rezultatus (RMSE 0,0910; δ_1 0,8580), aplenkdamas visus bazinius modelius. Nepaisant aukštų skaičių, vizualinis vertinimas išryškino šio modelio praktinius trūkumus: sugeneruoti gylio žemėlapiai yra labiau susilieję ir turi daugiau vizualinio triukšmo. Nors stikliniai kūnai juose yra sėkmingai atpažįstami ir paryškinami, už jų esantys fono objektai tampa neryškūs ir blankūs.

Dirbtinio intelekto įrankių naudojimas

Eksperimentiniais tyrimo tikslais pasitelktas „DepthAnythingV2“ gylio nustatymo dirbtinio intelekto modelis ir „Google Gemini 3 Pro Image“ dirbtinio intelekto modelis. Iliustracijai sugeneruoti naudotas „Google Gemini 3 Pro Image“ dirbtinio intelekto modelis (žr. **7 pav.**). Terminų vertimui ir teisingam programinio kodo formatavimui naudotas „Google Gemini 3 Pro“ modelis.

Literatūros sąrašas

1. ASKAR, C. - STERNBERG, H. Use of Smartphone Lidar Technology for Low-Cost 3D Building Documentation with iPhone 13 Pro: A Comparative Analysis of Mobile Scanning Applications. In *Geomatics* . 2023. Vol. 3, no. 4, p. 563–579.

2. HOTAIT, H. - FORRAI, A. An overview of monocular depth estimation with applicability in intelligent transportation. In *2025 IEEE 25th International Symposium on Computational Intelligence and Informatics (CINTI)* [interaktyvus]. Budapest, Hungary: IEEE, 2025. p. 000413–000418. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/11311818/>.

3. KHAN, F. ir kt. Deep Learning-Based Monocular Depth Estimation Methods—A State-of-the-Art Review. In *Sensors* . 2020. Vol. 20, no. 8, p. 2272.

4. YANG, L. ir kt. Depth Anything V2. In *NeurIPS 2024* [interaktyvus]. 2024. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2406.09414>.

5. PICCINELLI, L. ir kt. UniDepth: Universal Monocular Metric Depth Estimation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)* [interaktyvus]. [s.l.]: IEEE, 2024. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2403.18913>.

6. RANFTL, R. ir kt. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. In *IEEE TPAMI* [interaktyvus]. 2020. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1907.01341>.

7. VYAS, P. ir kt. Outdoor Monocular Depth Estimation: A Research Review. In *arXiv* [interaktyvus]. 2022. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://arxiv.org/abs/2205.01399>.

8. XU, H. ir kt. Seeing Glass: Joint Point Cloud and Depth Completion for Transparent Objects. In *arXiv* [interaktyvus]. 2021. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2110.00087>.

9. SAJJAN, S. ir kt. Clear Grasp: 3D Shape Estimation of Transparent Objects for Manipulation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* [interaktyvus]. Paris, France: IEEE, 2020. p. 3634–3642. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/9197518/>.

10. DAI, Q. ir kt. Domain Randomization-Enhanced Depth Simulation and Restoration for Perceiving and Grasping Specular and Transparent Objects. In *arXiv* [interaktyvus]. 2022. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2208.03792>.

11. WEN, H. ir kt. Seeing and Seeing Through the Glass: Real and Synthetic Data for Multi-Layer Depth Estimation. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2503.11633>.

12. TOMMASI, T. ir kt. A Deeper Look at Dataset Bias. In *arXiv* [interaktyvus]. 2015. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1505.01257>.

13. TOBIN, J. ir kt. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In *arXiv* [interaktyvus]. 2017. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1703.06907>.

14. TREMBLAY, J. ir kt. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. In *arXiv* [interaktyvus]. 2018. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1804.06516>.

15. MUMUNI, A. ir kt. A survey of synthetic data augmentation methods in computer vision. In *Machine Intelligence Research* . 2024. Vol. 21, no. 5, p. 831–869.

16. GUO, J. ir kt. UtilGen: Utility-Centric Generative Data Augmentation with Dual-Level Task Adaptation. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2510.24262>.

17. VALVANO, G. ir kt. Controllable Image Synthesis of Industrial Data Using Stable Diffusion. In *arXiv* [interaktyvus]. 2024. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2401.03152>.

18. WENG, Z. ir kt. Diffusion-HPC: Synthetic Data Generation for Human Mesh Recovery in Challenging Domains. In *arXiv* [interaktyvus]. 2023. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2303.09541>.

19. ATAPOUR-ABARGHOUEI, A. - BRECKON, T.P. Real-Time Monocular Depth Estimation Using Synthetic Data with Domain Adaptation via Image Style Transfer. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* [interaktyvus]. Salt Lake City, UT: IEEE, 2018. p. 2800–2810. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/8578394/>.

20. HENDRYCKS, D. ir kt. The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* [interaktyvus]. Montreal, QC, Canada: IEEE, 2021. p. 8320–8329. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/9710159/>.

21. TOSI, F. ir kt. Diffusion Models for Monocular Depth Estimation: Overcoming Challenging Conditions. In LEONARDIS, A. ir kt.Sud. *Computer Vision – ECCV 2024* [interaktyvus]. Cham: Springer Nature Switzerland, 2025. p. 236–257. [žiūrėta 2026-05-11]. ISBN 978-3-031-73336-9. Prieiga per internetą: <https://link.springer.com/10.1007/978-3-031-73337-6_14>.

22. CHEN, Y. ir kt. Guided Diffusion-based Generation of Adversarial Objects for Real-World Monocular Depth Estimation Attacks. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://arxiv.org/abs/2512.24111>.

23. GASPERINI, S. ir kt. R4Dyn: Exploring Radar for Self-Supervised Monocular Depth Estimation of Dynamic Scenes. In *2021 International Conference on 3D Vision (3DV)* [interaktyvus]. London, United Kingdom: IEEE, 2021. p. 751–760. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/9665968/>.

24. BEJERANO, E. ir kt. Sim2Radar: Toward Bridging the Radar Sim-to-Real Gap with VLM-Guided Scene Reconstruction. In *arXiv* [interaktyvus]. 2026. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://arxiv.org/abs/2602.13314>.

25. SAJJAN, S.S. ir kt. ClearGrasp: 3D Shape Estimation of Transparent Objects for Manipulation. In *arXiv* [interaktyvus]. 2019. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1910.02550>.

26. XU, X. ir kt. Towards Ambiguity-Free Spatial Foundation Model: Rethinking and Decoupling Depth Ambiguity. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://arxiv.org/abs/2503.06014>.

27. SONG, Z. ir kt. DepthMaster: Taming Diffusion Models for Monocular Depth Estimation. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://arxiv.org/abs/2501.02576>.

28. RAMIREZ, P.Z. ir kt. TRICKY 2025 Challenge on Monocular Depth from Images of Specular and Transparent Surfaces. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://ieeexplore.ieee.org/document/11375755>.

29. LI, K. ir kt. Physics-inspired self-learning framework for unsupervised depth estimation on non-Lambertian surfaces. In *Machine Learning: Science and Technology* . 2025. Vol. 6, no. 3, p. 035059.

30. HIGASHIUCHI, G. ir kt. Robust Self-Supervised Monocular Depth Estimation via Intrinsic Albedo-Guided Multi-Task Learning. In *Applied Sciences* . 2026. Vol. 16, no. 2, p. 714.

31. GUO, H. ir kt. Multi-view Reconstruction via SfM-guided Monocular Depth Estimation. In *CVPR 2025* [interaktyvus]. 2025. [žiūrėta 2026-05-11]. Prieiga per internetą: <https://arxiv.org/abs/2503.14483>.

32. DENG, H. ir kt. FuseGrasp: Radar-Camera Fusion for Robotic Grasping of Transparent Objects. In *IEEE Transactions on Mobile Computing* . 2025. Vol. 24, no. 8, p. 7028–7041.

33. OHARA, M. ir kt. The Role of Specular Reflections and Illumination in the Perception of Thickness in Solid Transparent Objects. In *Frontiers in Psychology* . 2022. Vol. 13, p. 766056.

34. BLAKE, A. - BÜLTHOFF, H. Does the brain know the physics of specular reflection? In *Nature* . 1990. Vol. 343, no. 6254, p. 165–168.

35. WEIBEL, J.-B. ir kt. Challenges of Depth Estimation for Transparent Objects. In *In Review* [interaktyvus]. 2024. [žiūrėta 2026-05-16]. Prieiga per internetą: <https://www.researchsquare.com/article/rs-4270684/v1>.

36. CHENG, Y. ir kt. Rethinking Transparent Object Grasping: Depth Completion with Monocular Depth Estimation and Instance Mask. In *IEEE Robotics and Automation Letters* . 2026. Vol. 11, no. 5, p. 5510–5517.

37. VASILJEVIC, I. ir kt. DIODE: A Dense Indoor and Outdoor DEpth Dataset. In *arXiv* [interaktyvus]. 2019. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/1908.00463>.

38. SILBERMAN, N. ir kt. Indoor Segmentation and Support Inference from RGBD Images. In FITZGIBBON, A. ir kt.Sud. *Computer Vision – ECCV 2012* [interaktyvus]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 746–760. [žiūrėta 2026-03-21]. ISBN 978-3-642-33714-7. Prieiga per internetą: <http://link.springer.com/10.1007/978-3-642-33715-4_54>.

39. MASOUMIAN, A. ir kt. Monocular Depth Estimation Using Deep Learning: A Review. In *Sensors* . 2022. Vol. 22, no. 14, p. 5353.

40. RAJAPAKSHA, U. ir kt. Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey. In *ACM Computing Surveys* . 2024. Vol. 56, no. 12, p. 1–51.

41. EIGEN, D. ir kt. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. In *arXiv* [interaktyvus]. 2014. [žiūrėta 2026-05-16]. Prieiga per internetą: <http://arxiv.org/abs/1406.2283>.

42. GODARD, C. ir kt. Digging Into Self-Supervised Monocular Depth Estimation. In *arXiv* [interaktyvus]. 2019. [žiūrėta 2026-05-16]. Prieiga per internetą: <http://arxiv.org/abs/1806.01260>.

43. GARIFULLIN, K. ir kt. MaterialFusion: High-Quality, Zero-Shot, and Controllable Material Transfer with Diffusion Models. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-05-14]. Prieiga per internetą: <http://arxiv.org/abs/2502.06606>.

44. HE, K. ir kt. Mask R-CNN. In *arXiv* [interaktyvus]. 2017. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://arxiv.org/abs/1703.06870>.

45. ROMBACH, R. ir kt. High-Resolution Image Synthesis with Latent Diffusion Models. In *arXiv* [interaktyvus]. 2022. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2112.10752>.

46. LI, Y. ir kt. Benchmarking Detection Transfer Learning with Vision Transformers. In *arXiv* [interaktyvus]. 2021. [žiūrėta 2026-05-14]. Prieiga per internetą: <https://arxiv.org/abs/2111.11429>.

47. Black Forest Labs FLUX.2 Klein [9b] Hugging face. In *Black Forest Labs FLUX.2 Klein [9b] Hugging face* [interaktyvus]. 2026. [žiūrėta 2026-05-16]. Prieiga per internetą: <https://huggingface.co/black-forest-labs/FLUX.2-klein-9B>.

48. TEAM, G. ir kt. Gemini: A Family of Highly Capable Multimodal Models. In *arXiv* [interaktyvus]. 2025. [žiūrėta 2026-03-21]. Prieiga per internetą: <http://arxiv.org/abs/2312.11805>.

49. Free picture: vine, glass, cup, white, table. In *Pixnio - Public Domain Images* [interaktyvus]. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://pixnio.com/objects/glass/vine-glass-cup-on-white-table>.

50. KICKWURM. How does one fix this scratch in my glass table? [interaktyvus]. 2019. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://www.reddit.com/r/howto/comments/cv224l/how_does_one_fix_this_scratch_in_my_glass_table/>.

51. Empty glass of water which sat empty until I finally asked to have refilled - Picture of MW Restaurant, Oahu - Tripadvisor. In *Tripadvisor* [interaktyvus]. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://www.tripadvisor.com/LocationPhotoDirectLink-g60982-d5287766-i269565491-MW_Restaurant-Honolulu_Oahu_Hawaii.html>.

52. 7 Ingenious Ways to Reuse Old Empty Bottles (You'll Never Throw Them Out Again!). In *Apartment Therapy* [interaktyvus]. 2024. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://www.apartmenttherapy.com/what-to-do-with-excess-glass-bottles-37393446>.

53. Empty Glass Ampoule at ₹ 6.2/piece | Chennai | ID: 23155159030. In *IndiaMart* [interaktyvus]. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://www.indiamart.com/proddetail/empty-glass-ampoule-23155159030.html>.

54. Buy GIFTBASH Empty Pauaa Quarter Clear Glass Bottle 180 ml with metal cap and plastic seal | Perfect for Storage Liquor+Coffee+Mocktails | Gifting and Party Favors | Pack of 3 | Online at Low Prices in India - Amazon.in. In *Amazon.in* [interaktyvus]. [žiūrėta 2026-05-21]. Prieiga per internetą: <https://www.amazon.in/GIFTBASH-Quarter-Bottle-plastic-Mocktails/dp/B0FQ4BR2KZ>.