



**VILNIUS
TECH**

Elektroninių sistemų
katedra

Vykintas VYŠNIAUSKAS

**AUTOMATINĖ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖ TAIKANT
GILIOJO MOKYMOSI METODUS**

**AUTOMATIC ANALYSIS OF AEROPHOTOGRAPHY IMAGES USING
DEEP LEARNING METHODS**

Magistro baigiamasis darbas

Informacinių elektroninių sistemų studijų programa, valstybinis kodas 6211BX015

Dirbtinio intelekto sistemų specializacija

Informatikos inžinerijos studijų kryptis

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
ELEKTRONIKOS FAKULTETAS
ELEKTRONINIŲ SISTEMŲ KATEDRA

TVIRTINU
Katedros vedėjas

(parašas)
Prof. dr. D. Navakauskas
2021 m. _____ mėn. ____ d.

Vykintas VYŠNIAUSKAS

**AUTOMATINĖ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖ
TAIKANT GILIOJO MOKYMO SI METODUS**
**AUTOMATIC ANALYSIS OF AEROPHOTOGRAPHY IMAGES USING
DEEP LEARNING METHODS**

Magistro baigiamasis darbas

Informacinių elektroninių sistemų studijų programa, valstybinis kodas 6211BX015

Dirbtinio intelekto sistemų specializacija

Informatikos inžinerijos studijų kryptis

Vadovas

doc. dr. Dalius Matuzevičius

(pedag. vardas, moksl. laipsnis, vardas, pavardė) (parašas) (data)

Lietuvių kalbos konsultantas

dr. Anželika Gaidienė

(pedag. vardas, moksl. laipsnis, vardas, pavardė) (parašas) (data)

Vilniaus Gedimino technikos universitetas
Elektronikos fakultetas
Elektroninių sistemų katedra

ISBN _____ ISSN _____
Egz. sk.
Data-.....-.....

Antrosios pakopos studijų **Informacinių elektroninių sistemų** programos magistro baigiamasis darbas

Pavadinimas **Automatinė aerofotografinių nuotraukų analizė taikant giliojo mokymosi metodus**

Autorius **Vykintas Vyšniauskas**

Vadovas **Dalius Matuzevičius**

Kalba: lietuvių

Anotacija

Šiame darbe buvo tiriamos automatinės aerofotografinių nuotraukų analizės, taikant giliojo mokymosi metodus, galimybės. Tikslui pasiekti buvo įvykdyti šie uždaviniai: atlikta automatinės aerofotografinių nuotraukų analizės metodų analitinė apžvalga. Atrinkti giliojo mokymosi metodai, perspektyvūs taikyti automatinei aerofotografinių nuotraukų analizei. Įgyvendinti ir eksperimentiškai ištirti modernūs segmentavimo modeliai, taikytini aerofotografinėms nuotraukoms segmentuoti. Dėl nagrinėjamų vaizdų sudėtingumo, objektų įvairovės ir kitų iššūkių, dabartiniams dirbtinio intelekto ir mašininio mokymosi metodams trūksta tikslumo, todėl nuotolinio aptikimo srityje vis dar reikia naudoti žmogaus intelektą.

Prasminiai žodžiai: gilusis mokymasis, sąsukos neuroniniai tinklai, kompiuterinė rega, nuotolinis aptikimas, vaizdų segmentavimas, aerofotografinė nuotrauka.

Vilnius Gediminas Technical University
Faculty of Electronics
Department of Electronic Systems

ISBN	ISSN
Copies No.	
Date-.....-.....	

Master Degree Studies **Information Electronics Systems** study programme Master Graduation Thesis

Title	Automatic Analysis of Aerophotography Images using Deep Learning Methods
-------	---

Author Vykintas Vyšniauskas

Academic supervisor	Dalius Matuzevičius
---------------------	----------------------------

Thesis language: Lithuanian

Annotation

This work reviews capabilities of deep learning-based methods for automatic analysis of aerial images. To reach the goal, the following tasks were completed: analytical review of methods for automatic analysis of aerial images. Selection of viable deep learning methods for automatic aerial images analysis. Experimental study and practical implementation of state-of-the-art segmentation models for aerial images segmentation. Due to complexity of aerial images, diversity of objects and other challenges, current methods of artificial intelligence and machine learning lack precision and therefore human intelligence still needs to be used in remote sensing applications.

Keywords: deep learning, convolutional neural networks, computer vision, remote sensing, image segmentation, aerial image.

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
ELEKTRONIKOS FAKULTETAS
ELEKTRONINIŲ SISTEMŲ KATEDRA

Technologijos mokslų sritis

Elektros ir elektronikos inžinerijos mokslo kryptis

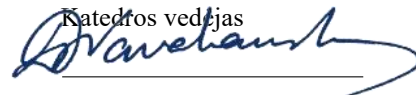
Informatikos inžinerijos studijų kryptis

Informacinių elektroninių sistemų studijų programa, valst. kodas 6211BX015

Dirbtinio intelekto sistemų specializacija

TVIRTINU

Katedros vedėjas



prof. dr. D. Navakauskas

2021 m. 04 mėn. 30 d.

BAIGIAMOJO MAGISTRO DARBO
UŽDUOTIS

Studentui

Vykintui VYŠNIAUSKUI

Baigiamojo darbo tema: Automatinė aerofotografinių nuotraukų analizė taikant giliojo mokymosi metodus

Automatic Analysis of Aerophotography Images Using Deep Learning Methods

Patvirtinta 2019 m. 11 mėn. 08 d. dekanų potvarkiu Nr. 233el

Baigiamojo darbo užbaigimo terminas 2021 m. 05 mėn. 20 d.

Darbo tikslas – ištirti automatinės aerofotografinių nuotraukų analizės taikant giliojo mokymosi metodus galimybes.

Darbo uždaviniai

1. Atlikti automatinės aerofotografinių nuotraukų analizės metodų analitinę apžvalgą.
2. Atrinkti giliojo mokymosi metodus, perspektyvius taikyti automatinei aerofotografinių nuotraukų analizei.
3. Įgyvendinti ir eksperimentiškai ištirti atrinktus automatinės aerofotografinių nuotraukų analizės metodus, ir pateikti tyrimo rezultatus.

Aiškinamojo rašto turinys

Įvadas

Darbo aktualumas ir tikslas

Darbo uždaviniai

Naudoti tyrimo ir analizės metodai

Darbo naujumas ir praktinė nauda

Darbo struktūra

1. Automatinės aerofotografinių nuotraukų analizės metodų analizės apžvalga

2. Atrinkti giliojo mokymosi metodai ir jų taikymas automatinei aerofotografinių nuotraukų analizei
3. Duomenų paruošimas automatinės aerofotografinių nuotraukų analizės metodų eksperimentiniam tyrimui
4. Giliojo mokymosi metodų, skirtų aerofotografinių nuotraukų analizei, tyrimo rezultatai

Apibendrinimas. Išvados

Literatūra

Priedai

A priedas. Automatinės aerofotografinių nuotraukų analizės metodų eksperimentinio tyrimo rezultatai

B priedas. Pranešimo XXIV-oje Lietuvos jaunųjų mokslininkų konferencijoje „MOKSLAS – LIETUVOS ATEITIS“ medžiaga

Vadovas

(parašas)

doc. dr. Dalius Matuzevičius

(pedag. vardas, moksl. laipsnis, vardas, pavardė)

Užduotį gavau

(parašas)

2021 m. balandžio mėn. 14 d.

(Baigiamojo darbo sąžiningumo deklaracijos forma)

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Vykintas Vyšniauskas, 20152381

(Studento vardas ir pavardė, studento pažymėjimo Nr.)

Elektronikos fakultetas

(Fakultetas)

Informacinės elektroninės sistemos, DISfm-19

(Studijų programa, akademinė grupė)

BAIGIAMOJO DARBO (PROJEKTO)

SĄŽININGUMO DEKLARACIJA

2021 m. gegužės 26 d.

Patvirtinu, kad mano baigiamasis darbas tema „Automatinė aerofotografinių nuotraukų analizė taikant giliojo mokymosi metodus“ patvirtintas 2019 m. lapkričio 08 d. dekanų potvarkiu Nr. 233el, yra savarankiškai parašytas. Šiame darbe pateikta medžiaga nėra plagijuota. Tiesiogiai ar netiesiogiai panaudotos kitų šaltinių citatos pažymėtos literatūros nuorodose.

Parenkant ir įvertinant medžiagą bei rengiant baigiamąjį darbą, mane konsultavo mokslininkai ir specialistai: docentas daktaras Dalius Matuzevičius. Mano darbo vadovas docentas daktaras Dalius Matuzevičius.

Kitų asmenų indėlio į parengtą baigiamąjį darbą nėra. Jokių įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs (-usi).



(Parašas)

Vykintas Vyšniauskas

(Vardas ir pavardė)

Automatinė aerofotografinių nuotraukų analizė taikant giliojo mokymosi metodus. Magistro baigiamasis darbas informatikos inžinerijos laipsniui. Vilniaus Gedimino technikos universitetas. Vilnius, 2021, 67 p., 51 iliustr., 24 lent., 77 bibl., 2 priedai.

Šiame darbe buvo tiriamos automatinės aerofotografinių nuotraukų analizės, taikant giliojo mokymosi metodus, galimybės. Tikslui pasiekti buvo įvykdyti šie uždaviniai: atlikta automatinės aerofotografinių nuotraukų analizės metodų analitinė apžvalga. Atrinkti giliojo mokymosi metodai, perspektyvūs taikyti automatinei aerofotografinių nuotraukų analizei. Įgyvendinti ir eksperimentiškai ištirti modernūs segmentavimo modeliai, taikytini aerofotografinėms nuotraukoms segmentuoti. Dėl nagrinėjamų vaizdų sudėtingumo, objektų įvairovės ir kitų iššūkių, dabartiniams dirbtinio intelekto ir mašininio mokymosi metodams trūksta tikslumo, todėl nuotolinio aptikimo srityje priimant sprendimus vis dar reikia naudoti žmogaus intelektą.

Raktiniai žodžiai: gilusis mokymasis, sąsūkos neuroniniai tinklai, kompiuterinė rega, nuotolinis aptikimas, vaizdų segmentavimas, aerofotografinė nuotrauka.

This work reviews capabilities of deep learning-based methods for automatic analysis of aerial images. To reach the goal, the following tasks were completed: analytical review of methods for automatic analysis of aerial images. Selection of viable deep learning methods for automatic aerial images analysis. Experimental study and practical implementation of state-of-the-art segmentation models for aerial images segmentation. Due to complexity of aerial images, diversity of objects and other challenges, current methods of artificial intelligence and machine learning lack precision and therefore human intelligence still needs to be used in remote sensing applications.

Keywords: deep learning, convolutional neural networks, computer vision, remote sensing, image segmentation, aerial image.

TURINYS

Įvadas	16
Darbo aktualumas ir tikslas.....	17
Darbo uždaviniai	17
Naudoti tyrimo ir analizės metodai	17
Darbo naujumas ir praktinė nauda	18
Darbo struktūra	18
1. AUTOMATINĖS AEROFOTGRAFINIŲ NUOTRAUKŲ ANALIZĖS METODŲ ANALITINĖ APŽVALGA	19
1.1. Dirbtinių neuronų tinklai.....	19
1.2. Gilieji neuroniniai tinklai	20
1.3. Sąsūkos neuroniniai tinklai	21
1.3.1. Aktyvacijos funkcijos.....	22
1.3.2. Sąsūkos sluoksnis	23
1.3.3. Sutelkimo sluoksnis.....	24
1.3.4. Visiškai sujungtų sąryšių sluoksnis	24
1.3.5. Klasifikavimo sluoksnis	25
1.4. Vaizdų analizės metodai	26
1.4.1. Segmentavimas, paremtas slenksčio pritaikymu	26
1.4.2. Segmentavimas, paremtas kraštinių aptikimu	28
1.4.3. Segmentavimas, paremtas vaizdo sričių panašumais	29
1.4.4. Segmentavimas, paremtas klasterizavimu	33
1.4.5. Segmentavimas, paremtas dirbtinių neuronų tinklais.....	34
1.4.6. Apibendrintas metodų palyginimas	34
1.5. Sąsūkos neuroninių tinklų architektūros	36
1.5.1. „VGGNet“	36
1.5.2. „ResNet“	36
1.5.3. „GoogLeNet“	37
1.5.4. „DenseNet“	38
1.6. Trumpas skyriaus apibendrinimas	38
2. ATRINKTI GILIOJO MOKYMOSI METODAI IR JŲ TAIKYMAS AUTOMATINEI AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZEI	39
2.1. Aeronuotraukų segmentavimo iššūkiai.....	39
2.2. Segmentavimo modeliai.....	42

2.2.1. VKT	42
2.2.2. „U-Net“	42
2.2.3. FPN	43
2.2.4. „PSPNet“	44
2.2.5. „LinkNet“	44
2.2.6. „DeepLabV3+“	46
2.3. Tikslų funkcijos	47
2.3.1. Dvejetainė kryžminė entropija	47
2.3.2. Pasvertoji dvejetainė kryžminė entropija	47
2.3.3. Balansuotoji kryžminė entropija	48
2.3.4. Židinio tikslų funkcija	48
2.3.5. „Dice“ tikslų funkcija	48
2.3.6. Tverskio tikslų funkcija	49
2.3.7. Tverskio židinio tikslų funkcija	49
2.3.8. Jautrumo ir specifiškumo tikslų funkcija	49
2.3.9. Jungtinė tikslų funkcija	49
2.3.10. Eksponentinė–logaritminė tikslų funkcija	50
2.4. Įvertinimo metrikos	51
2.4.1. Pikselių tikslumas (PA)	51
2.4.2. IoU	51
2.4.3. Preciziškumas, jautrumas ir „F1“ įvertis	52
2.4.4. „Dice“ koeficientas	52
2.5. Segmentavimo modelių tyrimas dvejetainio segmentavimo atveju	52
2.5.1. Kelių segmentavimas	53
2.5.2. Pastatų segmentavimas	58
2.6. Segmentavimo modelių tyrimas daugelio klasių segmentavimo atveju	61
2.7. Trumpas skyriaus apibendrinimas	65
3. DUOMENŲ PARUOŠIMAS AUTOMATINĖS AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖS METODŲ EKSPERIMENTINIAM TYRIMUI	67
3.1. Masačusetso kelių duomenų rinkinys	67
3.2. „DeepGlobe“ kelių duomenų rinkinys	68
3.3. Masačusetso pastatų duomenų rinkinys	69
3.4. „INRIA“ pastatų duomenų rinkinys	70
3.5. Lietuvos keliai	70
3.6. „DroneDeploy“ duomenų rinkinys	71

3.7. Duomenų papildymas	72
3.8. Trumpas skyriaus apibendrinimas	73
4. GILIOJO MOKYMOSI METODŲ, SKIRTŲ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZEI, TYRIMO REZULTATAI	74
4.1. Programinė įranga ir kompiuteriniai ištekliai	74
4.2. Dvejetainio segmentavimo rezultatai.....	75
4.2.1. Kelių segmentavimas.....	75
4.2.2. Pastatų segmentavimas	78
4.3. Daugelio klasių segmentavimo rezultatai	79
4.4. Trumpas skyriaus apibendrinimas	79
Apibendrinimas. Išvados	81
Literatūra	83
PRIEDAI	89
A priedas. Automatinės aerofotografinių nuotraukų analizės metodų eksperimentinio tyrimo rezultatai.....	89
B priedas. Pranešimo XXIV-oje Lietuvos jaunųjų mokslininkų konferencijoje „MOKSLAS – LIETUVOS ATEITIS“ medžiaga.....	97

SANTRUMPOS

DKE	dvejtainė kryžminė entropija (angl. <i>Binary Cross Entropy</i>)
DNT	dirbtinių neuronų tinklai (angl. <i>Artificial Neural Networks</i>)
DP	daugiasluoksnis perceptronas (angl. <i>Multi-Layer Perceptron</i>)
ELF	eksponentinė išlyginančioji funkcija (angl. <i>Exponential Linear Unit</i>)
FPN	požymių piramidės tinklas (angl. <i>Feature Pyramid Network</i>)
GP	grafikos procesorius (angl. <i>Graphics Processing Unit</i>)
GTT	gilaus „tikėjimo“ tinklas (angl. <i>Deep Belief Network</i>)
IEPS	išplėstasis erdvinis piramidės sutelkimas (angl. <i>Atrous Spatial Pyramid Pooling</i>)
IoU	aibių sankirtos ir sąjungos santykio vertė (angl. <i>Intersection Over Union</i>)
KE	kryžminė entropija (angl. <i>Cross Entropy</i>)
KEP	konteksto erdvinė priklausomybė (angl. <i>Contextual Pyramid Attention</i>)
KKNP	klaidingai klasifikuoti neigiami pavyzdžiai (angl. <i>False Negatives</i>)
KKTP	klaidingai klasifikuoti teigiami pavyzdžiai (angl. <i>False Positives</i>)
LF	išlyginančioji aktyvacijos funkcija (angl. <i>Rectified Linear Unit</i>)
LID	loginė išskirtinė disjunkcija (angl. <i>Exclusive OR</i>)
mIoU	vidutinė aibių sankirtos ir sąjungos santykio vertė (angl. <i>Mean Intersection Over Union</i>)
PA	pikselių tikslumo įvertis (angl. <i>Pixel Accuracy</i>)
SGD	tikimybinis gradientinio nusileidimo algoritmas (angl. <i>Stochastic Gradient Descent</i>)
SNT	sąsūkos neuroniniai tinklai (angl. <i>Convolutional Neural Networks</i>)
TKNP	teisingai klasifikuoti neigiami pavyzdžiai (angl. <i>True Negatives</i>)
TKTP	teisingai klasifikuoti teigiami pavyzdžiai (angl. <i>True Positives</i>)
VKT	Visiškai konvoliucinis tinklas (angl. <i>Fully Convolutional Network</i>)
VSS	visiškai sujungti sąryšiai (angl. <i>Fully Connected</i>)
ŽTF	židinio tikslo funkcija (angl. <i>Focal Loss</i>)

PAVEIKSLŲ SĄRAŠAS

1.1.1 pav. a) MP modelis ir b) vieno sluoksnio perceptronas [8, 9]	19
1.1.2 pav. DP modelis [10]	20
1.2.1 pav. Dirbtinio intelekto sritys [13]	21
1.3.1 pav. Klasikinė SNT architektūra [15]	21
1.3.2 pav. Aktyvacijos funkcijos: a) sigmoidė, b) LF, c) hiperbolinio tangento, d) ELF [18]	23
1.3.3 pav. Sąsūkos operacija tarp įvesties ir 3×3 dydžio sąsūkos filtro [19]	23
1.3.4 pav. Maksimalios reikšmės apibrėžtame lange išrinkimas [19]	24
1.3.5 pav. Matricos vertimas į vektorių [20]	24
1.3.6 pav. VSS sluoksnis [21]	25
1.3.7 pav. „Softmax“ aktyvacijos pavyzdys [23]	25
1.4.1 pav. Bendroji slenksčiu grindžiamo vaizdo segmentavimo struktūra [26]	27
1.4.2 pav. Slenksčio pritaikymo pavyzdžiai	28
1.4.3 pav. Pikselių plėtimo regione pavyzdys [32]	30
1.4.4 pav. Vaizdo dalinimas į regionus [25]	31
1.4.5 pav. Monochromatinis vaizdas [34]	32
1.4.6 pav. Topografinis vaizdas, kuriame matosi slėniai ir kalvos [34]	32
1.4.7 pav. Slėnių užtvindymas [34]	32
1.4.8 pav. Klasterizavimo pavyzdys [35]	34
1.5.1 pav. „VGG“ neuroninio tinklo architektūra [36]	36
1.5.2 pav. „ResNet“ neuroninio tinklo architektūra [37]	37
1.5.3 pav. „GoogLeNet“ neuroninio tinklo architektūra [39]	37
1.5.4 pav. „Inception“ blokas [39]	38
1.5.5 pav. „DenseNet“ neuroninio tinklo architektūra [42]	38
2.1.1 pav. Didelė tos pačios klasės objektų įvairovė [44]	40
2.1.2 pav. Didelis skirtingų klasių objektų panašumas [44]	40
2.1.3 pav. Didelis objekto ir scenos mastelių skirtumas [44]	41
2.1.4 pav. Kelių antžeminių objektų sambūvis [44]	41
2.2.1 pav. VKT architektūra [47]	42
2.2.2 pav. „U-Net“ architektūra [48]	43
2.2.3 pav. FPN architektūra [49]	44
2.2.4 pav. „PSPNet“ architektūra [50]	44
2.2.5 pav. „LinkNet“ architektūra [51]	45
2.2.6 pav. Enkoderio ir dekoderio blokų struktūra [51]	45

2.2.7 pav. IEPS blokas [52]	46
2.2.8 pav. „DeepLabV3+“ architektūra [53]	46
2.5.1 pav. Kelių segmentavimo, naudojant „DeepGlobe“ duomenų rinkinį, rezultatų vaizdavimas [5]	53
2.5.2 pav. Kelių segmentavimo, naudojant Inčuanos duomenų rinkinį, rezultatų vaizdavimas [56]	54
2.5.3 pav. Segmentavimo rezultatų palyginimas naudojant piramidės sutelkimo bloką (viduryje) ir nenaudojant jo (dešinėje) [58]	55
2.5.4 pav. Kelių segmentavimo, naudojant „PNOA“ duomenų rinkinį, rezultatų vaizdavimas [55]	56
2.5.5 pav. Kelių segmentavimo, naudojant Masačusetso kelių duomenų rinkinį, rezultatų vaizdavimas [60]	58
2.5.6 pav. Pastatų segmentavimo, naudojant Masačusetso pastatų duomenų rinkinį, rezultatų vaizdavimas [61]	59
2.5.7 pav. Pastatų segmentavimo, naudojant „INRIA“ duomenų rinkinį, rezultatų vaizdavimas. Viršuje yra originalus paveikslas ir jo tikroji kaukė, o apačioje modelio segmentavimo rezultatas be KEP bloko (kairėje) ir su KEP bloku (dešinėje) [63]	60
2.6.1 pav. 7 klasių segmentavimo, naudojant „DeepLab“ modelį ir „DeepGlobe“ duomenų rinkinį, rezultatų vaizdavimas [5]	62
2.6.2 pav. 20 klasių segmentavimo, naudojant „SkyScapes“ duomenų rinkinį, rezultatų vaizdavimas [65]	63
2.6.3 pav. Automobilių ir augalijos segmentavimo pavyzdys, taikant „DeepLabV3+“ modelį [66]	64
2.6.4 pav. 14 klasių segmentavimo, naudojant „iSAID“ duomenų rinkinį, rezultatų vaizdavimas [64]	65
3.1.1 pav. Masačusetso kelių duomenų rinkinio aerofotografinė nuotrauka ir jos kaukė [68]	68
3.2.1 pav. „DeepGlobe“ kelių duomenų rinkinio aerofotografinė nuotrauka ir kaukė [5]	69
3.3.1 pav. Masačusetso pastatų duomenų rinkinio aerofotografinė nuotrauka ir kaukė [68]	69
3.4.1 pav. „INRIA“ pastatų duomenų rinkinio aerofotografinė nuotrauka ir kaukė [69]	70
3.6.1 pav. „DroneDeploy“ duomenų rinkinio aerofotografinė nuotrauka ir kaukė [70]	71

LENTELIŲ SĄRAŠAS

1.4.1 lentelė. Vaizdo segmentavimo metodų palyginimas pagal [25] šaltinį	35
1.4.2 lentelė. Vaizdo segmentavimo metodų palyginimas pagal [26] šaltinį	35
2.3.1 lentelė. Tikslo funkcijų tipai [54]	47
2.3.2 lentelė. Įvairiems vaizdams segmentuoti naudojamų tikslo funkcijų apibendrinimas [54]	50
2.5.1 lentelė. Kelių segmentavimo rezultatai, naudojant Inčuanos kelių duomenų rinkinį [56]	54
2.5.2 lentelė. Kelių segmentavimo rezultatai, naudojant „DeepGlobe“ duomenų rinkinį [58]	55
2.5.3 lentelė. Kelių segmentavimo rezultatai, naudojant „PNOA“ duomenų rinkinį [55]	56
2.5.4 lentelė. Kelių segmentavimo rezultatai, naudojant Masačusetso kelių duomenų rinkinį [59]	57
2.5.5 lentelė. Kelių segmentavimo rezultatai, naudojant Masačusetso kelių duomenų rinkinį [60]	57
2.5.6 lentelė. Pastatų segmentavimo rezultatai, naudojant Masačusetso pastatų duomenų rinkinį [61]	59
2.5.7 lentelė. Pastatų segmentavimo rezultatai, naudojant „INRIA“ duomenų rinkinį [63]	60
2.5.8 lentelė. Pastatų segmentavimo rezultatai, naudojant Faihingenio prie Enco pastatų duomenų rinkinį [64]	61
2.5.9 lentelė. Pastatų segmentavimo rezultatai, naudojant Potsdamo pastatų duomenų rinkinį [64]	61
2.6.1 lentelė. 20 klasių segmentavimo rezultatai, naudojant „SkyScapes-Dense“ duomenų rinkinį [65]	63
2.6.2 lentelė. 6 klasių segmentavimo rezultatai, naudojant „DroneDeploy“ duomenų rinkinį [66]	63
2.6.3 lentelė. 14 klasių segmentavimo rezultatai, naudojant „iSAID“ duomenų rinkinį [64]	64
3.6.1 lentelė. Pikselių pasisikirstymas sužymėtų klasių „DroneDeploy“ duomenų rinkinyje [70]	71
3.7.1 lentelė. Pagrindinės duomenų papildymo technikos [72]	72
4.1.1 lentelė. GP techninių charakteristikų palyginimas	74
4.2.1 lentelė. Masačusetso kelių segmentavimo rezultatai, naudojant „Keras“ sistemą	75
4.2.2 lentelė. Masačusetso kelių segmentavimo rezultatai naudojant „PyTorch“ sistemą	76
4.2.3 lentelė. Lietuvos kelių segmentavimo rezultatai, naudojant „Keras“ sistemą	77
4.2.4 lentelė. „DeepGlobe“ kelių segmentavimo rezultatai, naudojant „PyTorch“ sistemą	77
4.2.5 lentelė. „INRIA“ pastatų segmentavimo rezultatai, naudojant „PyTorch“ sistemą	78

ĮVADAS

Vaizdas yra vienas iš esminių informacijos gavimo ir perdavimo būdų. Vaizdinės informacijos supratimas yra labai svarbus aspektas kompiuterinės regos (angl. *Computer Vision*) srityje, o pirmasis ir pagrindinis žingsnis analizuojant vaizdą – jo segmentavimas. Segmentavimas yra pagrįstas kriterijais, pagal kuriuos pradinis vaizdas padalijamas į skirtingas klases, siekiant išskirti tik dominančius objektus [1]. Vaizdo segmentavimas – individualių pikselių klasifikavimas (angl. *Semantic segmentation*) arba atskirų objektų iš pikselių klasės išskyrimas (angl. *Instance segmentation*) [2]. Atskirų objektų segmentavimas dar labiau išplečia semantinio segmentavimo aprėptį. Pavyzdžiui, nagrinėjant paveikslą, kuriame yra keturi asmenys, semantinio segmentavimo metu visi asmenys yra priskiriami vienai kategorijai, atskyrus juos nuo fono, o objektų segmentavimo metu iš tos kategorijos visi asmenys išskiriami kaip atskiri objektai. Daugelio klasių objektų segmentacija yra viena iš sudėtingiausių vaizdų analizės užduočių kompiuterinės regos srityje. Vaizdo padalijimas į atskirus segmentus yra svarbus medicinoje (aptinkant auglius ar kitus navikus), autonominių automobilių sistemose (aptinkant pėsčiuosius ar kitus objektus, lemiančius saugų eismą), robotikoje ir kitose srityse [3]. Iš esmės segmentavimo uždavinys nėra naujas, todėl literatūroje galima rasti daug skirtingų metodų, pradedant ankstyvaisiais klasikiniiais, pvz., slenksčiais, kraštinių aptikimu, sričių panašumais ar klasterizavimu grįštais segmentavimo metodais, baigiant moderniais giliojo mokymosi metodais. Pastarieji yra kur kas lankstesni ir našesni už klasikinius, nes pasiekia geriausius rezultatus tarp lyginamųjų indeksų tikslumo rodiklių. Pastaraisiais metais dirbtinių neuronų tinklai tapo viena labiausiai į mokslinius tyrimus orientuotų dirbtinio intelekto sričių. Dėl pasiektos pažangos kompiuterinės regos ir jutiklinių technologijų srityse, nuotolinio aptikimo (angl. *Remote sensing*) vaizdų segmentavimas ir objektų atpažinimas tampa vis aktualesnėmis mokslinių tyrimų temomis. Šiuo metu veikia daug operacinių palydovų, kuriuose įmontuoti įvairūs optiniai, mikrobangų, infraraudonųjų spindulių jutikliai, leidžiantys gauti visapusiškos informacijos apie žemės paviršių. Palydoviniai duomenys naudojami žvalgybai atlikti, miestams planuoti, augmenijai stebėti, gamtos ištekliams stebėti, klimato kaitos tyrimams atlikti, geologiniams pavojams stebėti, vandenynams stebėti ir daugelyje kitų sričių [4]. Visas taikymo sritis yra sunku aprėpti, todėl šiame darbe yra akcentuojami trys svarbūs aerofotografinių nuotraukų analizės uždaviniai – automatinis kelių, pastatų ir daugelio klasių segmentavimas, taikant giliojo mokymosi metodus.

Darbo aktualumas ir tikslas

Šiuolaikinių giliojo mokymosi metodų pritaikymas nuotolinio aptikimo duomenims analizuoti gali turėti didelės įtakos aplinkos supratimui ir lemti didesnius pasiekimus pasaulio miestų planavimo, klimato kaitos ir kitose mokslinių tyrimų srityse [3]. Aerofotografiniai vaizdai tik pastaruoju metu sulaukė didesnio dirbtinio intelekto bendruomenės dėmesio, sprendžiant tokias problemas kaip žemėlapių sudarymas, demografinių rodiklių analizė, efektyvus žemėtvarkos ir žemėnaudos planavimas, savaeigių transporto priemonių sistemų tobulinimas ir t. t. Miškai ir augmenija daro tiesioginę įtaką ekologijai ir biologinei įvairovei. Miškų plotų pokyčių stebėjimas leidžia planuoti miškų atsodinimą, vandens telkinių gaivinimą, taip pat ir natūralius gamtos pokyčius. Be to, galima užkirsti kelią plataus masto gaisrams ar įvertinti gaisrų padarytą žalą [3]. Aerofotografinės nuotraukos suteikia aiškią vietovės erdvinę informaciją, padedančios greitai ir lengvai atlikti įvairius skaičiavimus. Segmentuojant pastatus, galima lengviau atlikti 3D miestų rekonstrukciją. Išskiriant kelius, galima greičiau atnaujinti kelių tinklų žemėlapius. Nelaimių zonose, ypač besivystančiose šalyse, specifiniai žemėlapiai ir prieinamumo informacija yra labai svarbūs, norint greitai reaguoti į nelaimingus atsitikimus. Pirmasis reagavimo į krizes žingsnis gali būti automatinis kelių ir pastatų segmentavimas aerofotografinėse nuotraukose. Automatinis žemės dangos klasifikavimas ir segmentavimas yra labai svarbūs norint tvariai vystyti žemės ūkį, miškininkystę ir miestus, todėl keliamas platesnio masto uždavinys – daugelio klasių segmentavimas. Daug sudėtingų kompiuterinės regos užduočių, tokių kaip objektų aptikimas, lokalizacija, atpažinimas ir segmentavimas, yra veiksmingai sprendžiamos, taikant giliuosius dirbtinių neuronų tinklus. Tačiau dabartiniams metodams vis dar trūksta tikslumo [4]. Šio **darbo tikslas** – ištirti automatinės aerofotografinių nuotraukų analizės, taikant giliojo mokymosi metodus, galimybes.

Darbo uždaviniai

1. Atlikti automatinės aerofotografinių nuotraukų analizės metodų analitinę apžvalgą.
2. Atrinkti giliojo mokymosi metodus, perspektyvius taikyti automatinei aerofotografinių nuotraukų analizei.
3. Įgyvendinti ir eksperimentiškai ištirti atrinktus automatinės aerofotografinių nuotraukų analizės metodus ir pateikti tyrimo rezultatus.

Naudoti tyrimo ir analizės metodai

Šiame darbe buvo naudojamos nemokamos „Google Colab“ ir „Kaggle“ debesijų paslaugos. Šios platformos suteikia vartotojui kompiuterinius išteklius, kurie yra būtini eksperimentams atlikti. Buvo taikomos aukštojo lygio mašininio mokymo sistemos „Keras“ ir „PyTorch“. Taip pat buvo naudojamos šios pagrindinės „Python“ programavimo kalba parašytos bibliotekos:

- „segmentation-models“,
- „segmentation-models-pytorch“,
- „albumentations“,
- „Pytorch-toolbelt“.

Darbo naujumas ir praktinė nauda

Aerofotografiniai vaizdai yra turtingas ir struktūruotas informacijos šaltinis, tačiau kompiuterinės regos srityje jie yra tyrinėjami mažiau nei kasdieniai įprasti vaizdai. Dėl šių vaizdų sudėtingumo, objektų įvairovės, susiduriama su daugybe iššūkių, įskaitant daugiamatiškumą, neapibrėžtumą, netiesiškumą ir kt., todėl didžioji dalis dabartinių analizės metodų nėra visiškai tinkami [5]. Dirbtinio intelekto ir mašininio mokymosi metodams vis dar reikia naudoti žmogaus intelektą.

Darbo struktūra

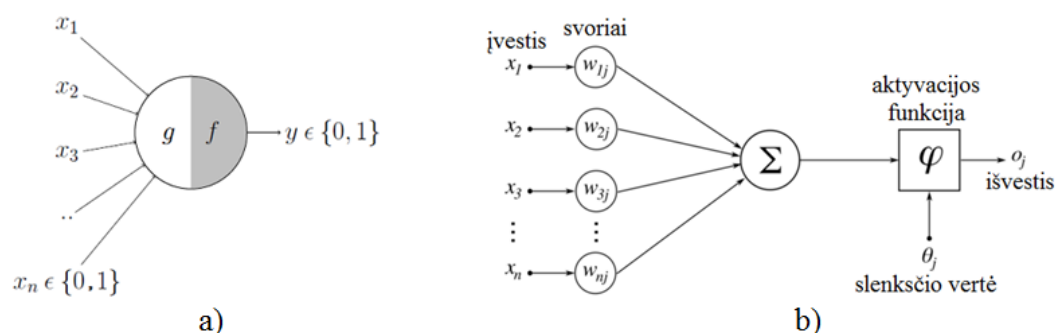
Šis darbas yra sudarytas iš keturių dalių. Pirmoje darbo dalyje aprašomi dirbtinių neuronų tinklai, klasikiniai ir giliojo mokymosi vaizdų analizės metodai. Pateikiamos pamatinės sąsūkos neuroninių tinklų architektūros. Antroje dalyje trumpai aprašoma, su kokiais iššūkiais susiduriama analizuojant aerofotografines nuotraukas. Nagrinėjami modernūs segmentavimo modeliai. Aprašomos pagrindinės segmentavimui skirtos tikslo funkcijos ir įvertinimo metrikos. Remiantis įvairiais tyrimais išrenkami giliojo mokymosi metodai, perspektyvūs taikyti automatinei aerofotografinių nuotraukų analizei. Trečioje darbo dalyje apibūdinami duomenų rinkiniai, kurie buvo panaudoti eksperimentams atlikti. Ketvirtoje darbo dalyje pateikiami atliktų eksperimentų rezultatai.

1. AUTOMATINĖS AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖS METODŲ ANALITINĖ APŽVALGA

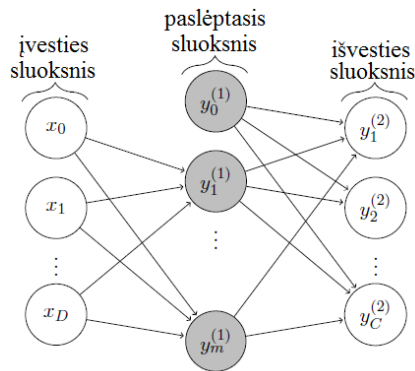
Šiame skyriuje aptariami dirbtinių neuronų tinklai ir nagrinėjama sąsukos neuroninių tinklų struktūra. Plačiau analizuojami ir lyginami klasikiniai vaizdų segmentavimo metodai. Nagrinėjamos pamatinės sąsukos neuroninių tinklų architektūros. Skyriaus pabaigoje pateikiamas trumpas skyriaus apibendrinimas.

1.1. Dirbtinių neuronų tinklai

1943 m. Varenas McCullochas ir Valteris Pitsas pasiūlė pirmąjį matematinį neuroninį modelį – MP modelį, kuris imituoja biologinių neuronų veikimą, atlikdamas tiesinius loginės algebros veiksmus ir sukurdamas dvejetainę išvestį, pagrįstą tam tikra slenksčio verte [6, 7]. MP dirbtinis neuronas yra pagrindinis DNT struktūrinis elementas, kuris gali būti padalintas į 2 dalis (žr. 1.1.1 (a) pav.). Pirmoji dalis g priima įvestis x ir atlieka veiksmus, o antroji dalis f , remdamasi apibendrinta verte, priima sprendimą [8]. Šeštojo dešimtmečio pabaigoje Frankas Rosenblattas pasiūlė vieno sluoksnio perceptrono (angl. *Single-Layer Perceptron*) modelį, papildydamas MP modelį mokymosi galimybėmis [6]. Perceptrone įvestims yra priskiriami tam tikri svorio koeficientai, kurie nusako kiekvienos įvesties svarbą, o aktyvacijos funkcija įveda į modelį netiesiškumus (žr. 1.1.1 (b) pav.). Perceptronas, kitaip nei MP modelis, gali atlikti veiksmus su realiaisiais skaičiais ir yra lankstesnis [8]. Apskritai, šie paprasčiausi modeliai yra geriau žinomi kaip dvejetainiai klasifikatoriai. Buvo aišku, kad perceptronas negali išspręsti tiesiškai neatskiriamų problemų, pavyzdžiui, LID operacijos, todėl buvo ieškota naujų sprendimų. 1986 m. buvo pasiūlyta (Hinton et al., 1986) daugiasluoksnio tiesioginio sklido (angl. *Feedforward*) neuroninio tinklo idėja. Taip pat DP buvo pritaikytas atgalinio sklido (angl. *Back Propagation*) algoritmas, kuris vienareikšmiškai apibūdina tinklo mokymąsi. Paprasčiausias daugiasluoksnis perceptronas yra sudarytas iš trijų sluoksnių – įvesties, paslėptojo ir išvesties sluoksnio (žr. 1.1.2 pav.). Dažniausiai kiekvienas dirbtinis neuronas turi ryšį su kito sluoksnio visais neuronais, todėl toks sluoksnis vadinamas VSS sluoksniu. Pagrindinis DP privalumas – gebėjimas spręsti netiesines problemas [6].



1.1.1 pav. a) MP modelis ir b) vieno sluoksnio perceptronas [8, 9]

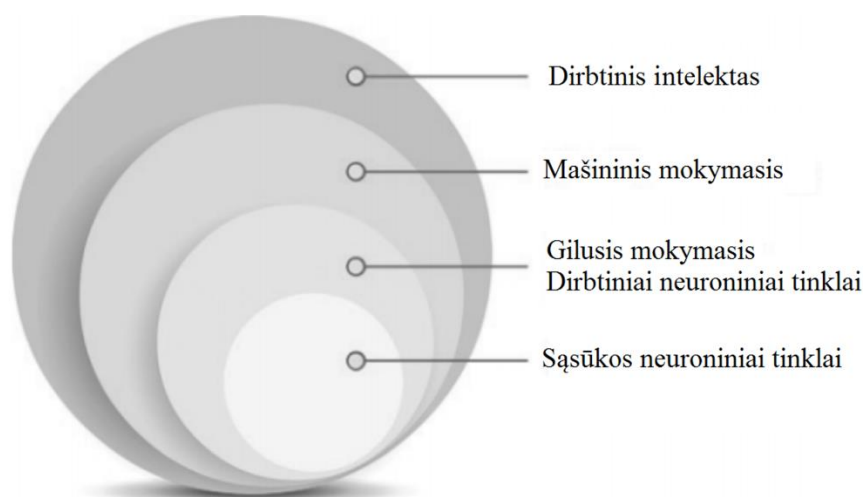


1.1.2 pav. DP modelis [10]

Dešimtajame dešimtmetyje vienas po kito buvo siūlomi įvairūs nesudėtingos struktūros neuroniniai tinklai, su vienu ar daugiau paslėptųjų sluoksnių, tačiau didesnio proveržio nebuvo pasiekta iki pat 2006 m., kai Hintonas su bendraautorais (Hinton et al., 2006) sukūrė gilaus „tikėjimo“ tinklą – GTT, kurio struktūra praktiškai nesiskiria nuo DP, bet mokymas vyksta pasluoksniui, naudojant godžiojo mokymo metodą. Autoriai įžvelgė tokių giliųjų „tikėjimo“ tinklų privalumus: a) jie neblogai geba išskirti reikšminius požymius, b) kiekvieno sluoksnio išankstinis apmokymas gali veiksmingai įveikti su mokymu susijusias problemas [6]. Šios įžvalgos padėjo pagrindą išsamesniems giliųjų neuroninių tinklų tyrimams ir atsirado nauja mašininio mokymosi sritis, vadinama giliuoju mokymusi. Vis dėlto, didinant paslėptųjų sluoksnių skaičių, tradiciniai atgalinio sklaidimo algoritmu grįsti tinklai susidurdavo su tokiomis problemomis kaip persimokymas, gradiento išnykimas, gradiento sprogimas, gradiento užstrigimas lokaliajame minimume ir kt. [6].

1.2. Gilieji neuroniniai tinklai

Įprastai giliaisiais vadinami tokie neuroniniai tinklai, kurie turi daugiau nei vieną paslėptąjį sluoksnį. Giliesiems neuroniniams tinklams būdingi sudėtingi algoritmai ir didelis duomenų kiekis [11, 12]. Kaip jau buvo minėta ankstesniame skirsnyje, tokie DNT turi akivaizdžių privalumų, lyginant su klasikiniiais, todėl sparčiai išpopuliarėjo. Giliojo mokymosi metodais siekiama sužinoti reikšminių požymių hierarchijas, pasižyminčias aukštesnio lygio savybėmis, kurias formuoja žemesnio lygio požymių sudėtis. Automatiškai išskirdamas požymius, giliojo neuroninio tinklo modelis gali išmokti sudėtingų funkcijų, susiejančių įvestį ir išvestį tiesiogiai iš duomenų. Tai ypač svarbu aukštesnio lygio abstrakcijoms, kurias žmonės dažnai nežino, kaip tiksliai apibrėžti, kalbant apie neapdorotą įvestį. Galimybė automatiškai išmokti sudėtingas funkcijas tampa vis svarbesnė, nes duomenų kiekis ir prieinamumas mašininio mokymosi metodams ir toliau auga. Nuo 2006 m. gilieji neuroniniai tinklai sėkmingai taikomi ne tik klasifikavimo užduotyse, bet ir regresijos, dimensių mažinimo, objektų segmentavimo, informacijos gavimo, robotikos, natūraliosios kalbos apdorojimo ir filtravimo srityse [12]. Gilusis mokymasis (angl. *Deep Learning*) yra mašininio mokymosi poaibis (žr. 1.2.1 pav.).

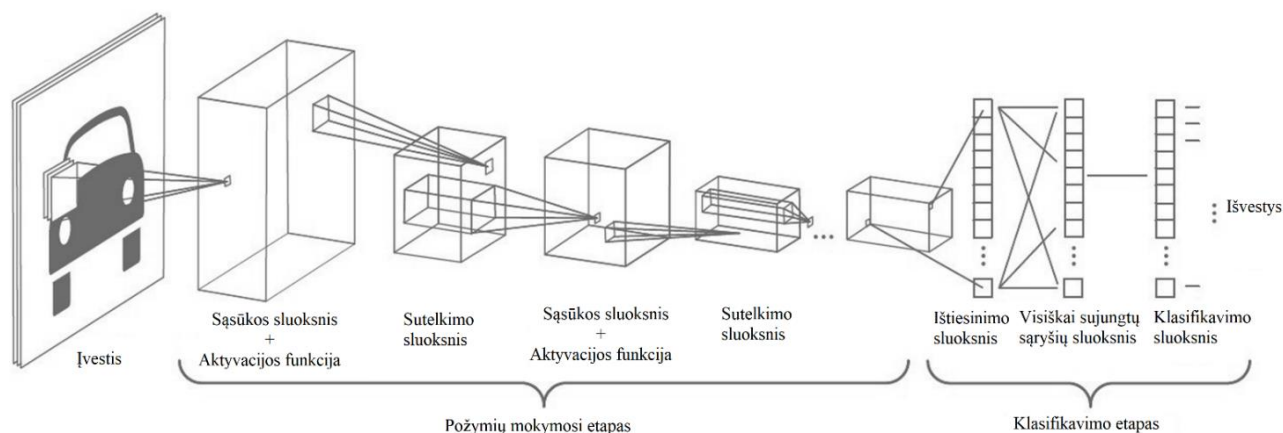


1.2.1 pav. Dirbtinio intelekto sritys [13]

Svarbu paminėti, kad buvo sukurta daug giliojo mokymosi metodų, iš kurių populiariausi yra sąsūkos, rekurentiniai ir generatyviniai priešiški neuroniniai tinklai. Šiame darbe bus nagrinėjami tik SNT, nes jie pasižymi puikiu gebėjimu analizuoti vaizdus. 2012 m. sukurtas sąsūkos neuroninis tinklas (Alex et al., 2012) pasiekė geriausias vaizdo klasifikavimo rezultatus „ImageNet Large Scale Visual Recognition Challenge“ varžytuvėse. Nuo to laiko stipriai išaugo SNT populiarumas ir atitinkamai didėjo tyrimų skaičius.

1.3. Sąsūkos neuroniniai tinklai

SNT pasiekė įspūdingų rezultatų kompiuterinės regos ir natūraliosios kalbos apdorojimo srityse, todėl pastaraisiais metais sulaukė didelio pramonės ir akademinės bendruomenės dėmesio [6]. Efektyvus giliųjų sąsūkos neuroninių tinklų mokymosi gebėjimas yra paremtas keliais požymių išskyrimo etapais, kuriuose iš pirminių duomenų yra automatiškai mokomasi reprezentacinių savybių. Didelio duomenų kiekio prieinamumas ir techninės įrangos technologijų pažanga paspartino SNT mokslinius tyrimus [14].



1.3.1 pav. Klasikinė SNT architektūra [15]

SNT yra savotiškas tiesioginio sklaidimo tinklas, kuris sugeba išskirti įvairius požymius iš duomenų, naudojant sąsūkos struktūras – filtrus (angl. *Kernels*). Paprasčiausią sąsūkos neuroninį tinklą sudaro

sąsūkos, sutelkimo ir visiškai sujungtų sąryšių sluoksniai (žr. 1.3.1 pav.). Tokią architektūrą įkvėpė biologinis vaizdų suvokimas [16]. Sąsūkos filtrai veikia kaip skirtingi receptoriai, galintys reaguoti į įvairius vaizdo požymius. Aktyvacijos funkcija užtikrina, kad informaciją kito sluoksnio neuronams perduos tik tie neuronai, kurių vertės viršija nustatytą ribą. Tikslo arba nuostolių funkcija apskaičiuoja skirtumą tarp tinklo išvesties vertės ir tikrosios, žmogaus nustatytos vertės. Tikslo funkcija yra labai svarbi neuroninio tinklo mokymosi priemonė. Tikslo funkcijai minimizuoti yra naudojami optimizavimo algoritmai (optimizatoriai) [6]. Populiariausi yra SGD, „Adam“ ir „RMSProp“ optimizavimo algoritmai. Lyginant su klasikiniiais dirbtiniais neuroniniais tinklais, SNT pasižymi tokiais privalumais:

- Vietinis ryšys. Kiekvienas neuronas jau nėra sujungtas su visais ankstesnio sluoksnio neuronais, o tik su nedideliu neuronų skaičiumi, kas efektyviai mažina parametrų skaičių ir spartina konvergenciją.
- Sviurių dalinimasis. Jungčių grupė gali turėti tą patį svorį, tai dar labiau sumažina parametrų skaičių.
- Dimensijų mažinimas. Sutelkimo sluoksnis naudoja vaizdo vietos koreliacijos su atvaizdo mažinimu principą, kuris gali sumažinti duomenų kiekį, išsaugant naudingą informaciją. Jis taip pat gali sumažinti parametrų skaičių, pašalindamas nereikšmingus požymius. Dėl šių trijų patrauklių savybių SNT tampa vienu iš reprezentatyviausių algoritmų giliojo mokymosi srityje.

1.3.1. Aktyvacijos funkcijos

Aktyvacijos funkcija yra būtina, nes be jos visas tinklas supaprastėja į tiesinę regresiją [17]. Klasikinės yra sigmoidės (angl. *Sigmoid*), hiperbolinio tangento (angl. *Tanh*), lyginamoji ir eksponentinė lyginamoji aktyvacijos funkcijos (žr. 1.3.2 pav.). Matematiškai, atitinkamu eiliškumu, šios funkcijos aprašomos taip:

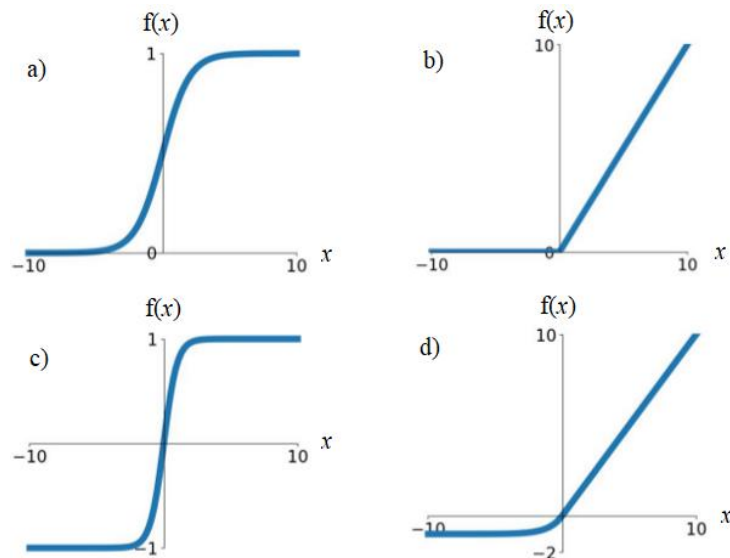
$$f(x) = \frac{1}{1 + e^{-x}}, \quad (1.3.1)$$

$$f(x) = \frac{e^x - 1}{e^x + 1}, \quad (1.3.2)$$

$$f(x) = \begin{cases} x, & \text{kai } x > 0 \\ 0, & \text{kai } x < 0 \end{cases}, \quad (1.3.3)$$

$$f(x) = \begin{cases} x, & \text{kai } x > 0 \\ \alpha(e^x - 1), & \text{kai } x < 0 \end{cases}, \quad (1.3.4)$$

čia x – įvesties tenzorius, α – parametras, nusakantis kreivės statumą.



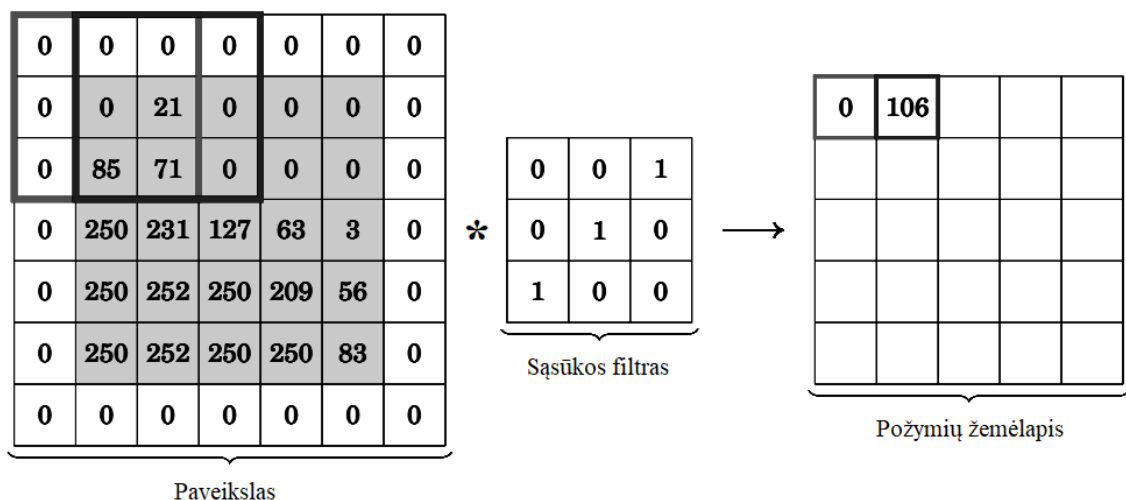
1.3.2 pav. Aktyvacijos funkcijos: a) sigmoidė, b) LF, c) hiperbolinio tangento, d) ELF [18]

1.3.2. Sąsūkos sluoksnis

Tai yra fundamentalusis SNT blokas, sudarytas iš nustatyto dydžio filtrų rinkinio. Filtras atlieka sąsūkos operaciją su įvestimi slenkančiojo lango principu, sudarydamas vadinamąjį požymių žemėlapi (žr. 1.3.3 pav.). Vaizdo dalijimas į mažus blokus leidžia gauti lokaliuosius požymius [14]. Šio sluoksnio parametrai yra filtrų svoriai, kurie mokymo metu keičiasi. Tame pačiame požymių žemėlapyje esantys neuronai dalijasi filtro svoriais, sumažindami tinklo sudėtingumą ir išlaikydami mažą parametrų skaičių [16]. Sąsūkos operacija matematiškai gali būti aprašoma taip:

$$f_l^k(p, q) = \sum_c \sum_{x, y} i_c(x, y) \cdot e_l^k(u, v), \quad (1.3.5)$$

čia $i_c(x, y)$ yra įvesties vaizdo tenzoriaus I_c elementas, kuris sluoksnyje l sudauginamas su k -ojo filtro elementu $e_l^k(u, v)$.



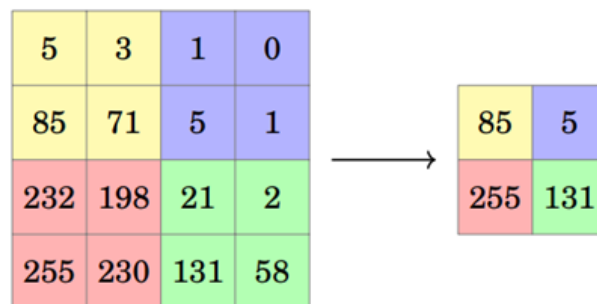
1.3.3 pav. Sąsūkos operacija tarp įvesties ir 3×3 dydžio sąsūkos filtro [19]

Požymių žemėlapiui paprastai pritaikoma netiesinė aktyvacijos funkcija. Sąsūkos neuroninis tinklas, atsižvelgiant į tinklo reikalavimus, gali susidėti iš skirtingo skaičiaus sąsūkos sluoksnių. Pirmieji

sluoksniai yra skirti aptikti žemo lygio vaizdo požymius, tokius kaip briaunos, kreivės, kampai, o aukštesniuose sluoksniuose esantys filtrai išmoka abstraktesnius požymius.

1.3.3. Sutelkimo sluoksnis

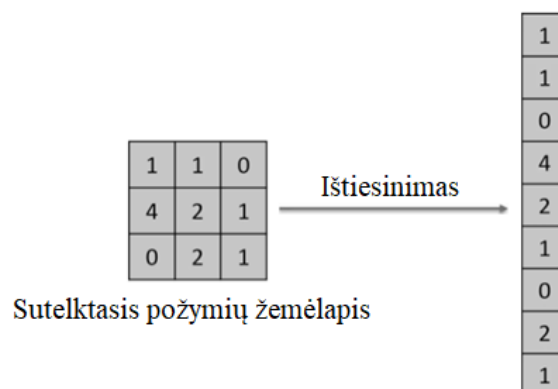
Duomenų sutelkimas yra svarbus sąsūkos neuroniniam tinklui, nes tai mažina jungčių tarp sąsūkos sluoksnių skaičių, todėl mažėja parametų skaičius. Sutelkimo sluoksniu (angl. *Pooling Layer*) siekiama padaryti poslinkio invariantiškumą, taip sumažinant požymių matricos dydį [14]. Sąsūkos neuroniniuose tinkluose toks sluoksnis paprastai yra tarp dviejų sąsūkos sluoksnių. Kiekvienas sutelkimo sluoksnio požymių žemėlapis yra sujungtas su atitinkamu, prieš tai esančio sluoksnio žemėlapiu. Tipinės sutelkimo operacijos yra vidutinės (angl. *Average Pooling*) arba maksimalios (angl. *Max Pooling*) reikšmės išrinkimas iš tam tikro dydžio duomenų matricos. Maksimalios reikšmės sutelkimo operacijos pavyzdys pateiktas 1.3.4 paveiksle.



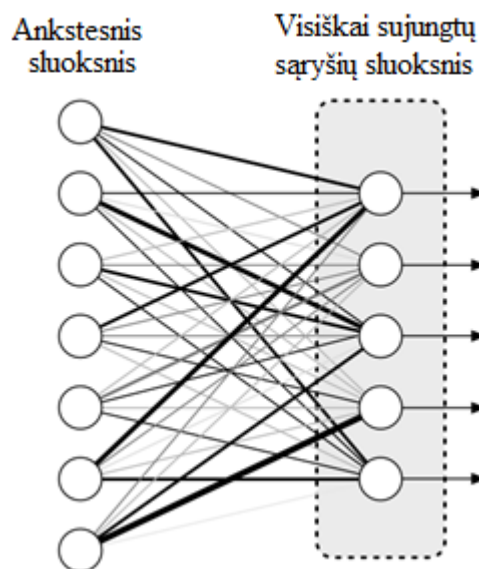
1.3.4 pav. Maksimalios reikšmės apibrėžtame lange išrinkimas [19]

1.3.4. Visiškai sujungtų sąryšių sluoksnis

Po kelių sąsūkos ir sutelkimo sluoksnių gali būti vienas ar daugiau VSS sluoksnių, kuriais siekiama atlikti aukšto lygio vaizdų klasifikavimą. Jie paima visus ankstesniame sluoksnyje esančius neuronus ir sujungia juos su kiekvienu dabartinio sluoksnio neuronu, sukurdami globalią semantinę vaizdo informaciją [6].



1.3.5 pav. Matricos vertimas į vektorį [20]

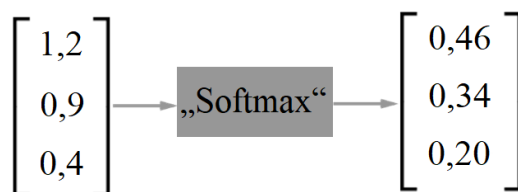


1.3.6 pav. VSS sluoksnis [21]

Viename iš visiškai sujungtų sąryšių sluoksnių paprastai visi požymių žemėlapiai (matricos) yra verčiami į vektorių (žr. 1.3.5 pav.). Po to eina dar vienas VSS sluoksnis, skirtas klasifikavimui. Kiekvienas šio sluoksnio neuronas yra prijungtas prie kiekvieno ankstesniame sluoksnyje esančio neurono (žr. 1.3.6 pav.). VSS sluoksnio tikslas yra paimti prieš tai esančio sluoksnio išvesties rezultatą ir priskirti jį tam tikrai iš anksto aprašytai klasei. VSS sluoksnis turi savo atgalinio sklaidimo mokymo algoritmą, kuris nustato reikiamus svorius. Kiekvienas neuronas gauna tam tikrą svorį, kurio vertė lemia galutinį tinklo rezultatą [15].

1.3.5. Klasifikavimo sluoksnis

Klasifikavimo arba nuostolių sluoksnis yra paskutinis VSS sluoksnis, kuris apskaičiuoja tinklo padarytas klaidas tarp tikrosios (pageidaujamos) išvesties ir tinklo išvesties. Klasifikavimo problemoms spręsti dažniausiai naudojama „Softmax“ aktyvacijos funkcija. Iš esmės tai yra multinominė regresija – aktyvacijos paverčiamos reikšmėmis tarp 0 ir 1, taip, kad jų suma būtų lygi 1. Didžiausia reikšmė rodo tinklo spėjimo tikimybę. Visiškai sujungtų sąryšių sluoksnis ne visada yra būtinas, nes jį galima pakeisti 1×1 filtro dydžio sąsūkos sluoksniu. Tokiu atveju tinklas vadinamas visiškai konvoliuciniu tinklu [22]. 1.3.7 paveiksle pateikiamas „Softmax“ funkcijos operacijos pavyzdys.



1.3.7 pav. „Softmax“ aktyvacijos pavyzdys [23]

1.4. Vaizdų analizės metodai

Vaizdo supratimas ir informacijos ištraukimas iš vaizdo tam tikroms užduotims atlikti yra svarbi skaitmeninio vaizdo technologijų taikymo sritis, o pirmas ir pagrindinis žingsnis suvokiant vaizdą yra jo segmentavimas. Per pastaruosius dešimtmečius buvo pasiūlyta daug vaizdo segmentavimo algoritmų. Apskritai, juos galima suskirstyti į penkias kategorijas [24]:

- Algoritmai, paremti slenksčio skaičiavimu (angl. *Threshold based*);
- Algoritmai, paremti kraštinių aptikimu (angl. *Edge based*);
- Algoritmai, paremti regionų radimu (angl. *Region based*);
- Algoritmai, paremti klasterizavimu (angl. *Clustering based*);
- Algoritmai, paremti dirbtinių neuronų tinklais (angl. *Neural Network based*).

Šiame skirsnyje apžvelgiami ir lyginami išvardintų kategorijų vaizdo segmentavimo algoritmai.

1.4.1. Segmentavimas, paremtas slenksčio pritaikymu

Slenksčio nustatymas – paprasčiausias vaizdų segmentavimo metodas. Pikseliai padalijami pagal jų intensyvumo lygius. Vertės, esančios žemiau arba aukščiau nustatytos ribos, gali būti atitinkamai klasifikuojamos kaip objektas arba fonas. Dažniausiai yra išskiriami trys bendri slenksčio nustatymo būdai [25]:

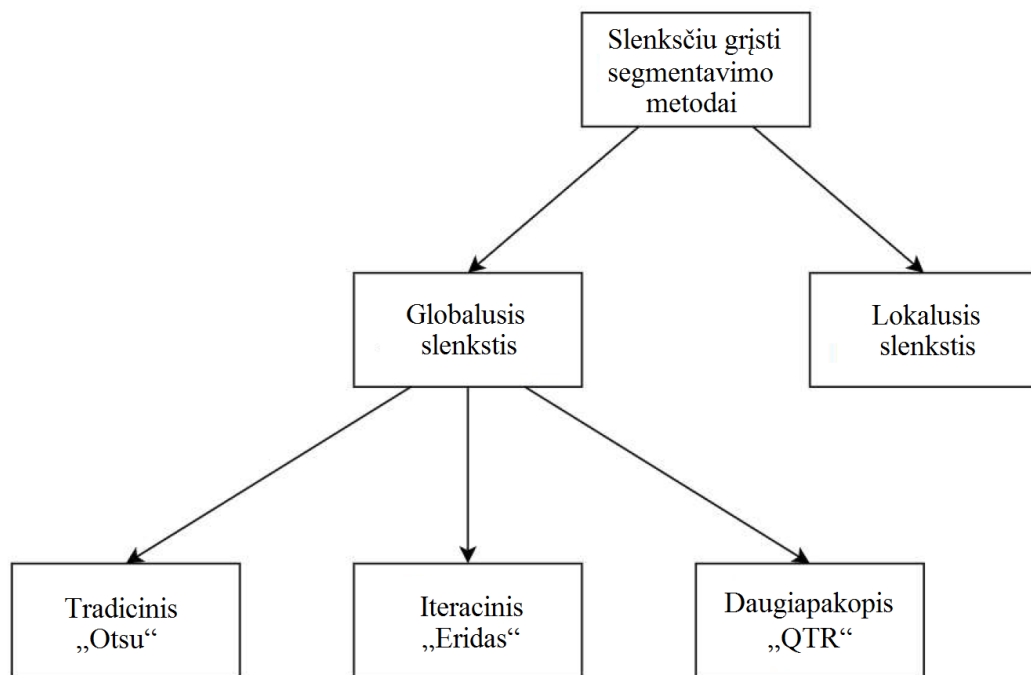
- Globalusis slenkstis. Segmentavimui naudojama tam tikra visam vaizdai pastovi slenkstinė vertė T . Pavyzdžiui, pradinio vaizdo takšai aprašomi taip: $p(x, y)$. Tuomet apdoroto vaizdo reikšmės $q(x, y)$ gaunamos taip:

$$q(x, y) = \begin{cases} 1, & \text{jeigu } p(x, y) > T \\ 0, & \text{jeigu } p(x, y) \leq T \end{cases} \quad (1.4.1)$$

- Kintamas slenkstis. Šiuo atveju slenksčio vertė T gali kisti, analizuojant skirtingus vaizdo taškus. Slenksčio vertė priklauso nuo gretimų tašku x ir y (lokalusis slenkstis) arba slenksčio vertė yra x ir y funkcija (adaptyvusis slenkstis).
- Kelių lygių slenkstis. Vaizdai segmentuoti naudojamos kelios slenksčio vertės T_0 ir T_1 , su kuriomis apdorotas vaizdas gali būti aprašytas taip:

$$q(x, y) = \begin{cases} m, & \text{jeigu } p(x, y) > T_1 \\ n, & \text{jeigu } p(x, y) \leq T_1 \\ o, & \text{jeigu } p(x, y) \leq T_0 \end{cases} \quad (1.4.2)$$

Kituose šaltiniuose pastarieji du būdai sujungiami į vieną, kaip parodyta 1.4.1 paveiksle [26].



1.4.1 pav. Bendroji slenksčiu grindžiamo vaizdo segmentavimo struktūra [26]

Slenksčio nustatymas iš histogramos. Slenksčio operacijos įvestys paprastai yra monochromatiniai arba spalvotieji vaizdai, o išvestys yra binariniai vaizdai, kuriuose balta spalva žymi objektą, o juoda – foną (arba atvirkščiai) [27]. Histograma grįsti metodai priklauso nuo to, ar pavyks įvertinti ribinę vertę, skiriančią objektą nuo fono. Pavyzdžiui, P_1 ir P_2 yra monochromatinio vaizdo histogramos pikai. Slenksčio vertė T apskaičiuojama taip: $T = (P_1 + P_2) / 2$. T gali būti pasirinkta kaip minimali vertė, esanti tarp tų pikų ($T = \min(H(u))$, kur $u \in [P_1, P_2]$, o $H(u)$ – histogramos vertė). Vaizdas segmentuojamas visą paveikslą dalinant į objektus ir foną, priklausomai nuo individualaus pikselio intensyvumo reikšmės ir slenksčio vertės T . Reikia pabrėžti, kad dažnai histogramos yra labai sudėtingos – turi daug viršūnių ir neaiškių slėnių, todėl ne visada lengva pasirinkti slenksčio vertę [28].

Iteracinis slenksčio nustatymas. Iteraciniai metodai duoda geresnį rezultatą, kai dėl triukšmų yra pakankamai sunku nustatyti slenkstį. Iteracinis metodas yra atsparesnis triukšmams. Algoritmo esmė yra tokia:

- vaizdo histograma pagal vidutinę vertę yra suskirstoma į dvi dalis (objektą ir foną),
- tuose dalyse taip pat randamos vidutinės vertės ir toliau nagrinėjama histogramos sritis, kuri jau yra tarp tų dviejų išrinktų verčių,
- nagrinėjamoje histogramos srityje vėl randama vidutinė vertė ir nustatomas naujas slenkstis.

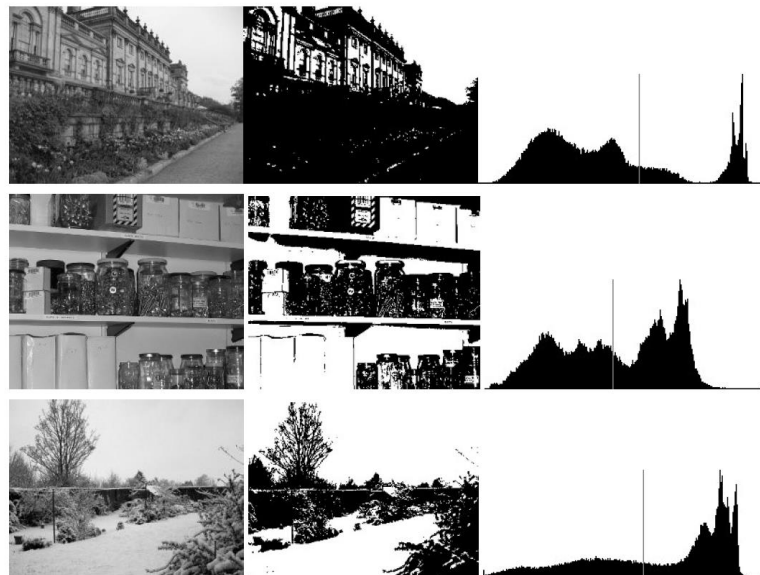
Pastarasis žingsnis vykdomas, tol kol nauja slenksčio vertė nebekinta [28]. k -toji slenksčio vertė apskaičiuojama taip:

$$T_{k+1} = \frac{\sum_{b=0}^{T_k} bn(b)}{2 \sum_{b=0}^{T_k} n(b)} + \frac{\sum_{b=T_{k+1}}^N bn(b)}{2 \sum_{b=T_{k+1}}^N n(b)}, \quad (1.4.3)$$

čia T_k yra slenksčio vertė k -oje iteracijoje, b yra intensyvumo vertė, $n(b)$ yra b reikšmės pasitaikymų skaičius. Šio metodo trūkumas yra tas, kad kaskart skaičiuojama nauja vidutinė vertė ir apskritai atliekama labai daug skaičiavimo operacijų.

„Otsu“ metodas. Šis slenksčio nustatymo metodas yra iteracinis ir apima visas galimas slenksčio vertes. Pagal kiekvieną slenkstį, vaizdas atskiriamas į dvi klases (objektą ir foną) ir apskaičiuojamos atskirų grupių histogramų vidutinės vertės. Žinant visų gautų slenksčių vertes, toliau pasirenkamos tokios, kurių tarpklasė dispersija yra didžiausia arba mažiausia. Iš esmės šis metodas tinka tik bimodales histogramas turintiems vaizdams [28]. Bimodalinėmis vadinamos tokios histogramos, kuriose aiškiai matosi du pikai.

Kelių lygių slenkstis. Globalaus slenksčio metodai gali būti išplečiami naudojant daugiau nei vieną slenkstį. Tokiu būdu galima vaizdą atskirti į kelias klases, pvz., fonas ir keli objektai. Vaizdas dalinamas pagal kelis intensyvumo lygius. Šis metodas tinka vaizdams, kurių fonas yra spalvotas ir nevientisas [27].



1.4.2 pav. Slenksčio pritaikymo pavyzdžiai

1.4.2. Segmentavimas, paremtas kraštinių aptikimu

Kalbant apie kraštinių aptikimo metodus pirmiausiai reikia aptarti, kas yra apibrėžiama kaip kraštinė vaizde [29]:

- kraštinė yra riba tarp dviejų sričių, pasižyminčių skirtingais intensyvumo lygiais, pvz., juoda–balta,
- kraštinės – pikseliai, kuriuose ryškio funkcija staigiai pasikeičia,
- kraštinių aptikimas – staigių vaizdo intensyvumo funkcijos pokyčių nustatymas.

Kraštinių aptikimo metodai yra gerai išvystyti. Jie grindžiami sparčiu intensyvumo vertės pokyčiu. Kraštinės randamos tose vaizdo dalyse, kuriose pirmoji intensyvumo išvestinė yra didesnė už slenkstį, arba antrosios išvestinės nulinio kirtimo vietose. Aptiktos kraštinės yra sujungiamos, suformuodamos

objekto ribas reikiama regionams segmentuoti. Apskritai yra išskiriami 2 kraštų aptikimo būdai [30]: gradiento – nustatomas funkcijos pirmosios išvestinės maksimumas arba minimumas ir Laplaso – nustatomos nulinio kirtimų vietos iš funkcijos antrosios išvestinės.

Gauso laplasianas. Šis algoritmas naudoja Gauso filtrą vaizdo suliejimui ir Laplaso operatorių (laplasianą) vaizdui gryninti. Pagrindiniai algoritmo žingsniai [30]:

- pritaikomas Gauso filtras (triukšmo mažinimas),
- skaičiuojamas laplasianas,
- surandami nulinio kirtimai,
- nulinio kirtimų diagramoje nustatomas staigus nuolydis arba pakylimas,
- staigaus pasikeitimo vietoje pritaikomas slenksčio metodas ir pažymimi kraštai.

Didžiausias šio metodo trūkumas yra aštrių kampų iškraipymas.

Kenio (angl. *Canny*) algoritmas. Tai dar vienas kraštinių aptikimo metodas, kurį sudaro tokie pagrindiniai žingsniai [30]:

- pritaikomas Gauso filtras,
- skaičiuojamas gradientas,
- vertės, kurios yra tarp maksimumų yra prislopinamos,
- pritaikomas daugiapakopio slenksčio principas.

Sobelio (angl. *Sobel*) operatorius. Apskaičiuojamas vaizdo intensyvumo gradientas kiekviename vaizdo pikselėje. Randamas sparčiausias intensyvumo pasikeitimas. Rezultatas parodo, kaip staigiai vaizdas keičiasi kiekviename pikselėje ir ar tas pikselis yra kraštinis. Tai taip pat rodo, kaip tas kraštas gali būti orientuotas. Kitaip sakant, gaunamas vektorius, nukreiptas per kraštinę nuo tamsesnių iki ryškesnių intensyvumo reikšmių. Sobelio filtras naudoja du 3×3 dydžio filtrus G_x ir G_y [31]:

$$G_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \cdot A, \quad G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \cdot A, \quad (1.4.4)$$

vienas – horizontalios krypties pokyčiams nustatyti, kitas – vertikalios krypties pokyčiams nustatyti. Atlikus sąsūkos operaciją tarp filtrų ir pradinio vaizdo, apskaičiuojamos išvestinių aproksimacijos. Kaip matyti, filtrų reikšmės yra teigiamos ir neigiamos, vadinasi apdoroto vaizdo neigiamos vertės rodo tamsias sritis, o teigiamos šviesias [31].

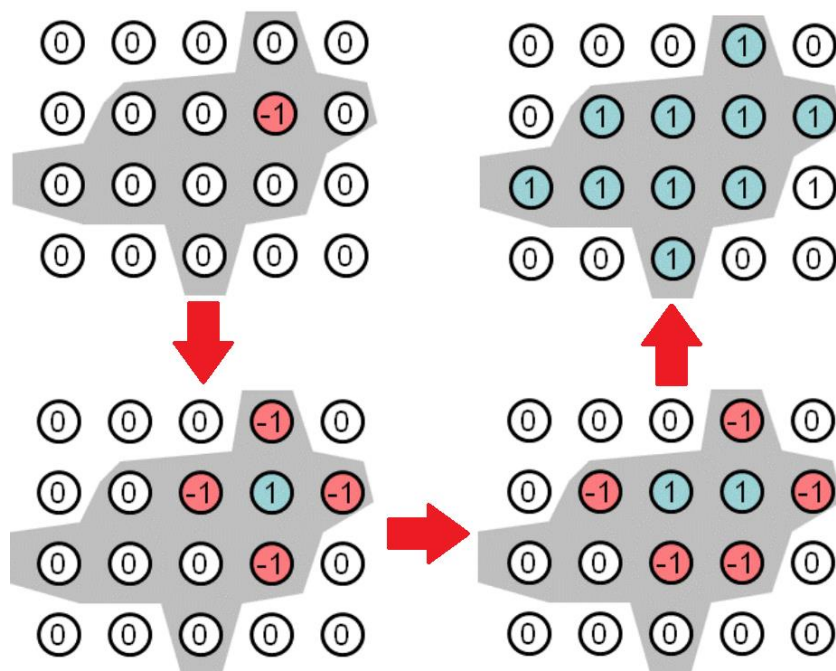
1.4.3. Segmentavimas, paremtas vaizdo sričių panašumais

Sričių panašumais grįsti segmentavimo metodai yra tokie, pagal kuriuos vaizdas padalijamas į įvairius regionus, pasižyminčius panašiomis savybėmis. Šiuo metodu paremti du pagrindiniai algoritmai: regiono augimo (angl. *Region Growing*), padalijimo ir suliejimo (angl. *Region Splitting And Merging*) [32].

Regiono augimo algoritmas. Plečiant pradinis individualius pikselius, vaizdas suskirstomas į įvairius regionus (žr. 1.4.3 pav.). Pradiniai pikseliai pasirenkami remiantis ankstesnėmis žiniomis arba pagal konkretų pritaikymą. Paprastas regiono plėtimo algoritmas yra paremtas 8 kaimyninių pikselių principu (angl. *8-Connectivity*). Pavyzdžiui, $p(x,y)$ yra pradinis paveikslas, kurį reikia segmentuoti, o $s(x,y)$ yra dvejetainis paveikslas, kuriame sudaryti plėtimo pikseliai [25, 32]. Tuomet, galima išskirti tokius pagrindinius algoritmo žingsnius:

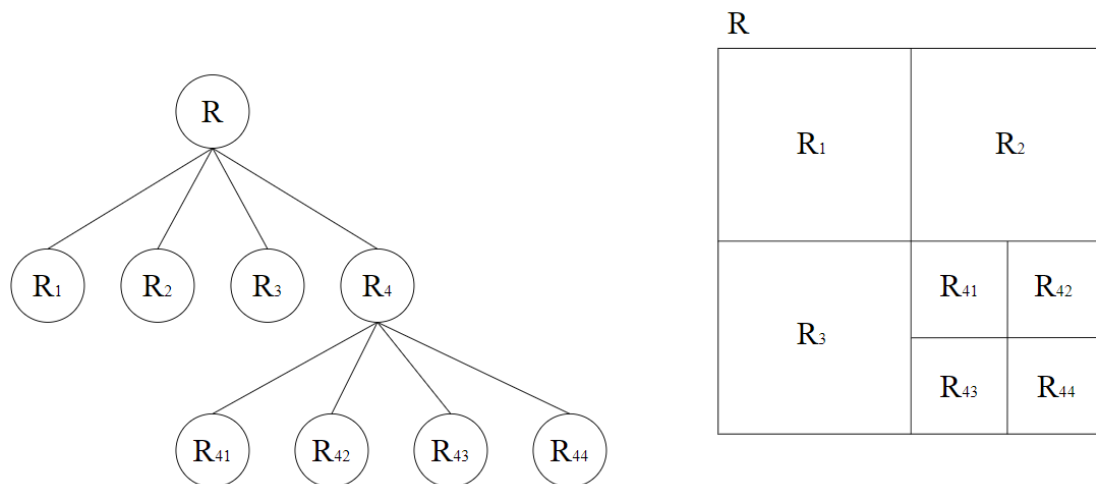
- pirmiausia, visiems sujungtiems $s(x,y)$ paveikslo komponentams atliekama morfologinė erozijos (angl. *Erosion*) operacija (kiekvienas individualus pikselis, kuris bus plečiamas, yra atskiriamas),
- apskaičiuojamas dvejetainis paveikslas P_T , kur $P_T(x,y) = 1$, jeigu tame taške yra plėtimo pikselis,
- apskaičiuojamas dvejetainis paveikslas q , kur $q(x,y) = 1$, jeigu $P_T(x,y) = 1$ ir tas taškas (x,y) yra sujungtas su $s(x,y)$ 8 kaimyninių pikselių principu.

Taigi, paveiksle q sujungti komponentai yra susegmentuoti regionai.



1.4.3 pav. Pikselių plėtimo regione pavyzdys [32]

Regionų padalijimo ir suliejimo algoritmas. Vaizdas dalinamas į regionus, kurie pasižymi panašiomis savybėmis, o panašūs regionai yra suliejami. Taip sudaromos įvairaus dydžio sritys. Įvesties duomenys grupuojami į regionų piramidinę struktūrą. 2D atveju kiekvienas regionas sugrupuotas į keturias grupes, o 3D atveju – į aštuonias grupes. (žr. 1.4.4 pav.) [25, 32].



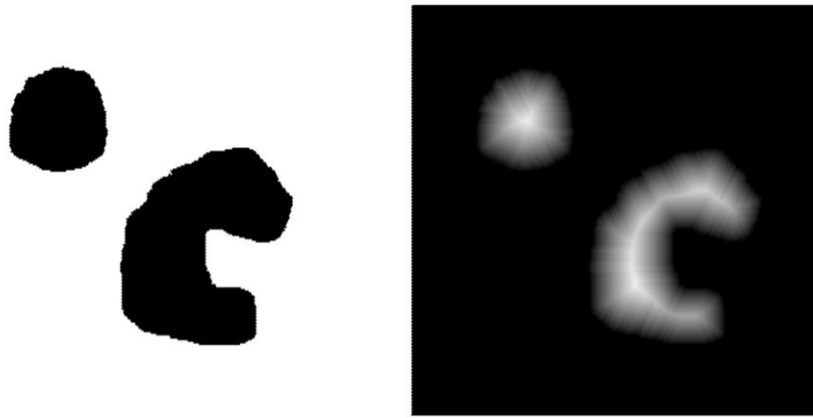
1.4.4 pav. Vaizdo dalinimas į regionus [25]

Pavyzdžiui, p yra pradinis vaizdas. Pagrindiniai regiono padalijimo ir suliejimo algoritmo etapai yra tokie [10]:

- pirmiausia, regionas R yra lygus p ,
- jeigu toje srityje skiriasi vaizdo elementų savybės, tai ta sritis yra dalijama į kvadrantus,
- jeigu atskirose srityse visi vaizdo elementai yra panašių savybių, tai jie sujungiami į bendrą regioną,
- dalijimas ir sujungimas vykdomas tol, kol galima išskirti panašių požymių sritis.

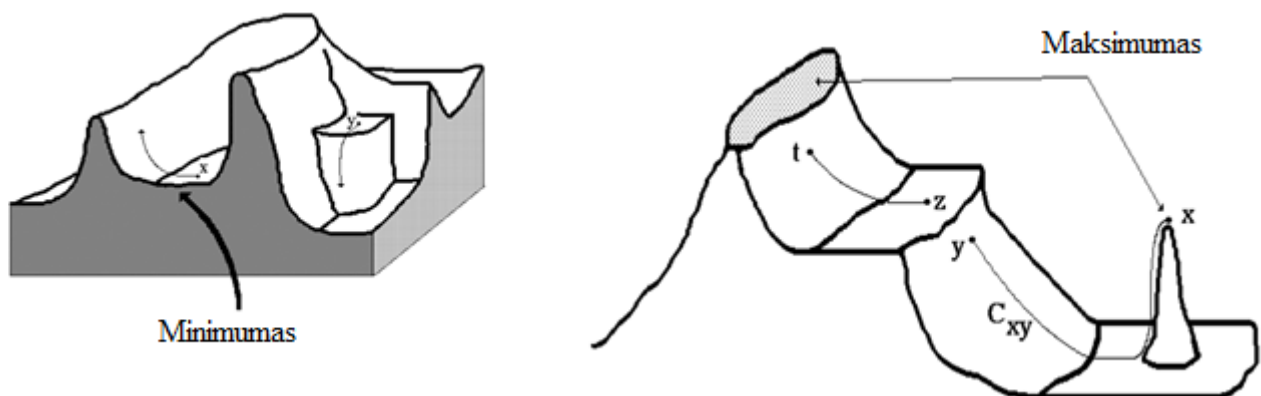
Šis algoritmas turi kelis pranašumus prieš kitus vaizdo segmentavimo algoritmus. Regiono augimo metodas yra paprastas ir labai stabilus triukšmo atžvilgiu [32].

Vandenskyros transformacija. Vandenskyros algoritmas yra plačiai naudojamas vaizdų segmentavimo metodas. Vienas iš pastebimų šio segmentavimo metodų trūkumų yra pernelyg didelė segmentacija. Spręsdami šią problemą, Mejeris ir Biučeris [33] pasiūlė vandenskyros transformacijos modifikaciją, kurioje segmentavimui pasitelkiamos objektų žymės. Pagrindinis šio sprendimo tikslas – pažymėti dominantį objektą vadinamuoju „objekto žymekliu“, o foną pažymėti „foniniu žymekliu“. Metodo principas yra nustatyti vaizdo gradientą ir pritaikyti vandenskyros transformaciją. Vietoje gradiento gali būti naudojamos kitos morfologinės operacijos. Gradientą galima apibūdinti kaip skirtumą tarp morfologinių išsiplėtimo ir erozijos operacijų. Paprastai šiam algoritmui yra naudojami monochromatiniai vaizdai. Daroma prielaida, kad funkcija, apibrėžianti monochromatinį vaizdą, yra tolydi. Tada gradientas lygus funkcijos moduliui. Modulis gali būti aproksimuotas, priskiriant individualiam pikseliui tam tikrą vertę, reiškiančią didelio ir mažo intensyvumo skirtumą kaimyniniuose pikseliuose.



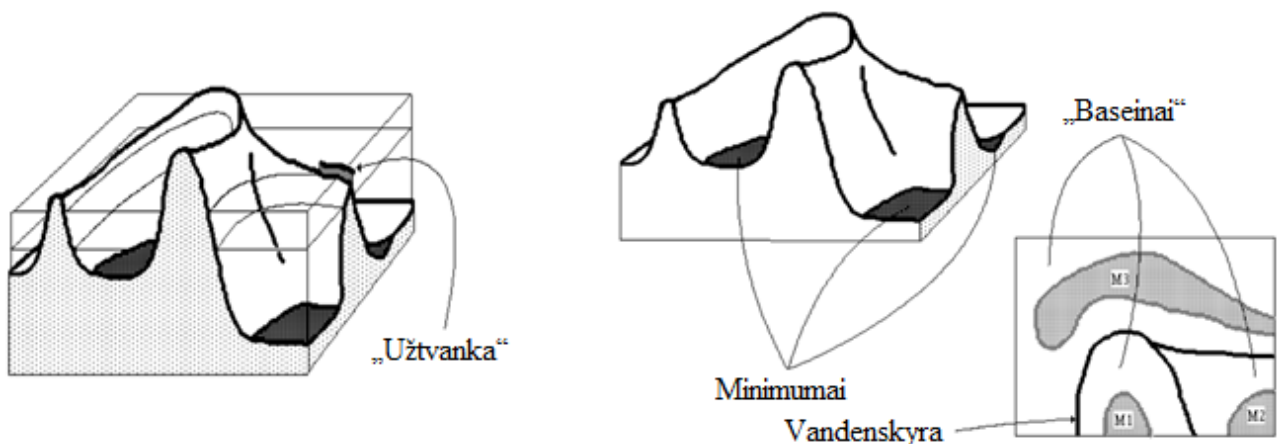
1.4.5 pav. Monochromatinis vaizdas [34]

Nagrinėjant monochromatinius vaizdus galima lengvai išskirti šviesesnes ir tamsesnes sritis, kurios apibūdinamos pagal intensyvumą. Dėl šios priežasties vaizdas gali būti atvaizduotas kaip topografinis paviršius, su minimumais ir maksimumais. Minimumai – tai duobės topografiniame vaizde, o maksimumai – kalvos.



1.4.6 pav. Topografinis vaizdas, kuriame matosi slėniai ir kalvos [34]

Kraštinių išskyrimas vykdomas, taikant imituojamą „potvynio“ procesą. Kiekviename minimume keliamas vandens lygis, užtvindant visą paviršių. Šio proceso metu užtvindytos vietos iš skirtingų minimumų susijungia. Susijungimo vietose sudaromi barjerai, atskiriantys objektus nuo fono [33].



1.4.7 pav. Slėnių užtvindymas [34]

Dėl tam tikrų triukšmų ir nuolydžių svyravimų, susidaro daug baseinų, su įvairiais vietiniais minimumais, todėl atliekant vandenskyros transformaciją, vaizdas segmentuojamas labai tankiai. To

galima išvengti, filtruojant staigius nuolydžius ir išlaikant tik svarbiausius, tačiau tai yra labai sudėtinga užduotis [33]. Dėl šios priežasties naudojama objektų žymekliais pagrįsta sistema. Žymekliai gali būti kuriami rankiniu būdu, sudarant vietinius minimumus objektui ir fonui, dėl ko gaunamas naujas gradientas. Fonas turi būti žymimas taip, kad apsuptų dominantį objektą. Naudojant naują gradientą, baseinai kuriami pagal pasirinktų žymeklių skaičių, segmentuojant vaizdą tik pasirinktose srityse.

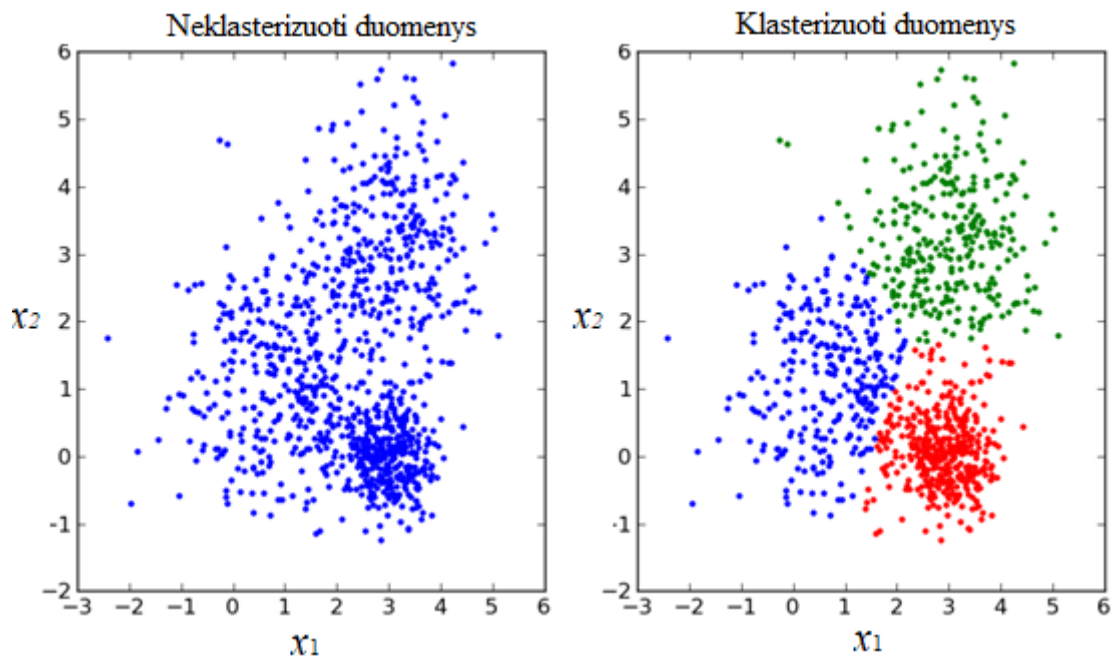
1.4.4. Segmentavimas, paremtas klasterizavimu

Klasterizavimo algoritmai segmentuoja panašių charakteristikų pikselius į klasterius. Yra dvi pagrindinės klasterizavimo metodų kategorijos: hierarchinis ir dalybos metodas. Hierarchinis metodas paremtas medžių samprata, kur medžio šaknis simbolizuoja visą vaizdą, o medžio šakos – klasterius. Dalyba grįsti metodai iteratyviai naudoja optimizavimo metodus, minimizuojančius tikslo funkciją. Dažniausiai išskiriami kietojo klasterizavimo (angl. *Hard Clustering*) ir minkštojo klasterizavimo (angl. *Soft Clustering*) algoritmai [25].

Kietojo klasterizavimo algoritmas. Tai paprastas grupavimo metodas, padalijantis vaizdą į klasterius taip, kad vienas pikselis galėtų priklausyti tik vienam klasteriui. Kiekvienam pikseliui naudojama funkcija, kurios reikšmės gali būti tik 1 arba 0, todėl pikselis priklauso klasteriui arba ne. Kaip pavyzdį galima paminėti „k-means“ klasterizavimo algoritmą. Pirmiausia skaičiuojami atskiro klasterio centrai visame paveiksle, tada kiekvienas pikselis priskiriamas artimiausiam centrui. Pagrindiniai „k-means“ algoritmo žingsniai yra šie [25]:

- atsitiktinai pasirinkami pirminiai klasterizavimo centrai,
- skaičiuojamas atstumas tarp kiekvieno pikselio ir kiekvieno klasterio centro. Pikselis, kurio atstumas iki tam tikro klasterio centro yra mažiausias, bus priskirtas tam klasteriui,
- kiekvienam klasteriui sudaroma nauja centro pozicija, pagal tą klasterį sudarančių pikselių atstumo iki centro vidurkį,
- veiksmai kartojami tol, kol klasterio centras nebesikeis arba bus pasiektas nustatytas iteracijų skaičius [25].

Minkštojo klasterizavimo algoritmas. Tai natūralusis klasterizavimo metodas, kuriame naudojamas negriežtas dalinimas į klasterius. Kitaip tariant, labai tiksliai klasterizuoti negalima, dėl triukšmų. Kaip pavyzdį galima paminėti neaiškia logika (angl. *Fuzzy Logic*) grįsta „c-means“ algoritmą, kur individualus pikselis gali priklausyti keliems klasteriams [25].



1.4.8 pav. Klasterizavimo pavyzdys [35]

„*k-means*“ klasterizavimo algoritmo privalumas yra tas, kad algoritmas yra greitas ir paprastas, o dideliems duomenų rinkiniams jis yra labai efektyvus ir išplečiamas. Jo sudėtingumas yra beveik tiesinis. Kita vertus, klasterio numeris K neturi aiškių atrankos kriterijų, todėl jį sunku įvertinti [25]. Taigi, „*k-means*“ taikomas, turint tokius duomenis, kurių funkcija yra iškilioji (angl. *Convex*). Klasterizavimo pavyzdys pateiktas 1.4.8 paveiksle.

1.4.5. Segmentavimas, paremtas dirbtinių neuronų tinklais

Dirbtinių neuronų tinklai tai šiuolaikiškas ir vis populiarėjantis metodas, kuris labai išplėtė vaizdų analizės galimybes. Lyginant su klasikineis segmentavimo metodais, DNT metodai yra lankstesni ir rodo geresnius rezultatus. Be to, DNT metodai gali būti automatizuoti. Šiuo metu egzistuoja daugiau nei 100 įvairių giliuoju mokymusi grįstų vaizdo segmentavimo metodų. Pabrėžtina, kad skirtingi metodai paprastai tinka tik specifinėms problemoms spręsti. Pavyzdžiui, modelis, kuris puikiai segmentuoja pastatus, nebūtinai gerai segmentuos kitus objektus. Apie giliojo mokymosi metodus plačiau aprašoma 2-ajame skyriuje.

1.4.6. Apibendrintas metodų palyginimas

Iš esmės, kiekvienas segmentavimo metodas turi savo privalumų ir trūkumų. Akivaizdu, kad skirtingi metodai tinka specifinėms taikymo situacijoms. Pavyzdžiui, paprasčiausi klasikiniai metodai labiau tinka tais atvejais, kai reikia atskirti vienai klasei priklausančius objektus nuo fono. Daugelio klasių segmentavimo atveju yra taikomi klasterizavimo ir dirbtinių neuronų tinklų metodai. Remiantis [25, 26], buvo sudarytos lentelės, kuriuose buvo palyginti šiame skyriuje aptarti vaizdų segmentavimo metodai (žr. 1.4.1–1.4.2 lenteles). Klasikiniai segmentavimo metodai nėra automatizuoti.

1.4.1 lentelė. Vaizdo segmentavimo metodų palyginimas pagal [25] šaltinį

Metodas	Aprašymas	Privalumai	Trūkumai
Slenksčio skaičiavimo	Slenksčiui nustatyti, remiamasi paveikslėlio histogramos viršūnėmis	Paprasčiausias metodas, kuriam be histogramos kitos informacijos nereikia	Labai priklauso nuo histogramos piku.
Kraštinių aptikimo	Segmentavimas atliekamas remiantis vaizdo netolygumais	Puikiai tinka vaizdams kurių kontrastas tarp objektų ir fono yra didelis	Netinkamas, kai kraštų yra labai daug
Regionų radimo	Segmentavimas atliekamas remiantis vaizdo skirstymu į homogenines sritis	Atsparus triukšmams. Naudingas tuomet, kai lengva išskirti požymius	Brangus laiko ir kompiuterio atminties atžvilgiu
Klasterizavimo	Segmentavimas atliekamas remiantis vaizdo dalinimu į homogeninius klasterius	Tinka natūralių vaizdų segmentavimui, nes pikselis gali priklausyti keliems klasteriams	Klasterizavimo funkcija yra sudėtinga
Neuroninių tinklų	Paremtas tinklo mokymosi ir sprendimų priėmimo principu	Nereikia rašyti sudėtingų programų	Apmokymas užima labai daug laiko.

1.4.2 lentelė. Vaizdo segmentavimo metodų palyginimas pagal [26] šaltinį

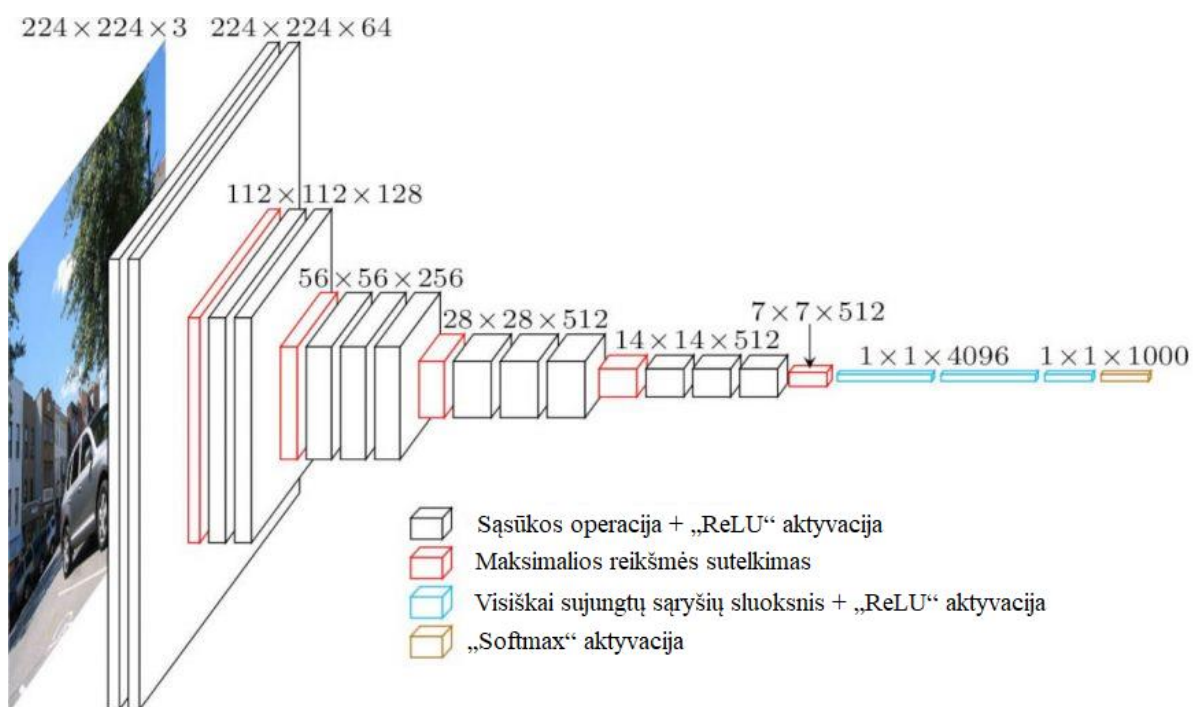
Metodas	Aprašymas	Privalumai	Trūkumai
Slenksčio skaičiavimo	Priklauso nuo histogramos	Paprastas metodas, kurio taikymui nereikia papildomų žinių	Neefektyvus skaičiavimo požiūriu. Netinka realaus laiko sistemoms
Kraštinių aptikimo	Segmentavimas atliekamas „nuo galo iki galo“ principu, nustatant ribas.	Išlaiko pilkus atspalvius kraštuose ir gerą kontrastą	Sunku pritaikyti mažo kontrasto vaizdams Netinka, jeigu kraštinių yra labai daug
Regionų radimo	Pagal Morfologines operacijas objektus atskiria į skirtingus regionus	Paprasti skaičiavimai ir greitos operacijos. Puikiai tinka didelio kontrasto vaizdams	Kartais regionai persidengia arba yra neteisingai nustatomi
Klasterizavimo	Pikselių, turinčių panašias savybes, grupavimas ir grupavimo reikšmės apibrėžiamos pagal jų intensyvumą	Puikiai tinka mažiems duomenų rinkiniams analizuoti. Duoda pakankamai gerą rezultatą	Ilgas ir brangus skaičiavimų laikas. Kadangi paremtas atstumų skaičiavimų, tai daug kur netinka
Neuroninių tinklų	Naudojami neuroninio tinklo mazgai.	Pranašesnis už visus kitus klasikinius metodus	Apmokymas užima labai daug laiko. Kartais galimas persimokymas

1.5. Sąsukos neuroninių tinklų architektūros

Šiame skirsnyje pateikiamos pamatinės SNT architektūros, kurios istoriškai užėmė aukščiausias vietas „ImageNet Large Scale Visual Recognition Challenge – ILSVRC“ vaizdų klasifikavimo varžytuvėse. „ImageNet“ yra duomenų rinkinys, kurį sudaro daugiau kaip 15 mln. sužymėtų aukštos raiškos vaizdų. Duomenų rinkinyje yra išskirta apie 22 000 objektų klasių [36]. „ILSVRC“ varžybų tikslas – skatinti tobulesnių kompiuterinės regos technologijų kūrimą ir palyginti naujausius metodus.

1.5.1. „VGGNet“

„VGGNet“ yra klasikinė sąsukos neuroninių tinklų architektūra, pasižyminti savo paprastumu. Tinkle yra naudojami 3×3 dydžio sąsukos filtrai, LF aktyvacijos funkcija, 2×2 dydžio maksimalios reikšmės sutelkimo filtrai, su žingsniu lygiu 2. Po sąsukos ir sutelkimo sluoksnių seka trys VSS sluoksniai: pirmuosiuose dviejuose yra po 4096 kanalų, o trečiame – 1000 kanalų (po vieną kiekvienai klasei). Paskutinis sluoksnis yra „Softmax“ klasifikavimo sluoksnis. Pažymima, kad tinklo įvesties paveikslas dydis yra fiksuotas 224×224 pikselių raiškos. 1.5.1 paveiksle pateikta „VGG16“ architektūra. Modelį sudaro 16 sąsukos sluoksnių, o bendras parametrų skaičius yra 138 mln. „VGG16“ 2014 m. „ILSVRC2014“ konkurse užėmė antrąją vietą. „VGG“ architektūros tinklai išlieka populiariūs, dėl savo griežtos struktūros ir efektyvumo [36].

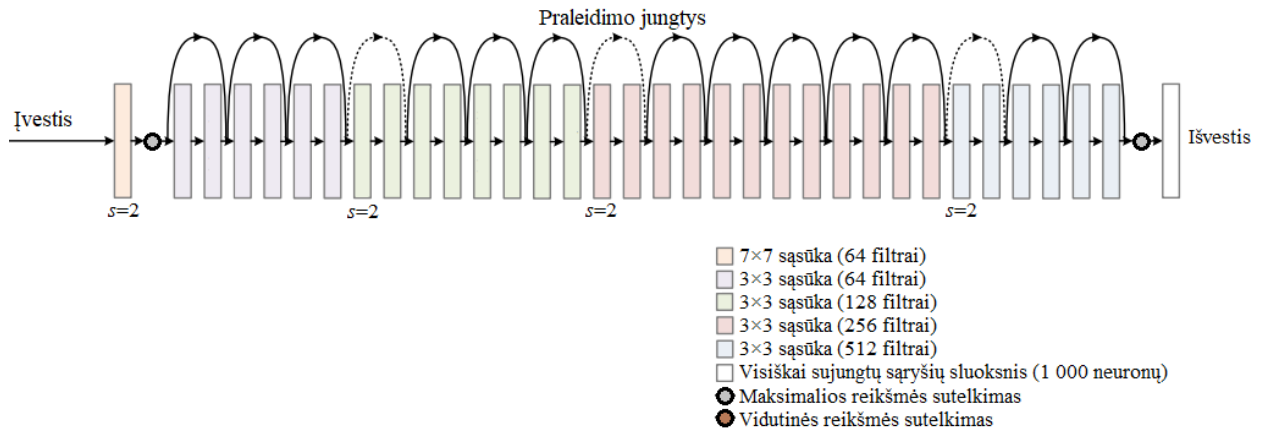


1.5.1 pav. „VGG“ neuroninio tinklo architektūra [36]

1.5.2. „ResNet“

Pagrindinis „ResNet“ architektūros elementas yra liekamųjų ryšių blokas (angl. *Residual Block*), kuriame yra vadinamosios praleidimo jungtys (angl. *Skip Connections*). Šios jungtys leido išspėsti

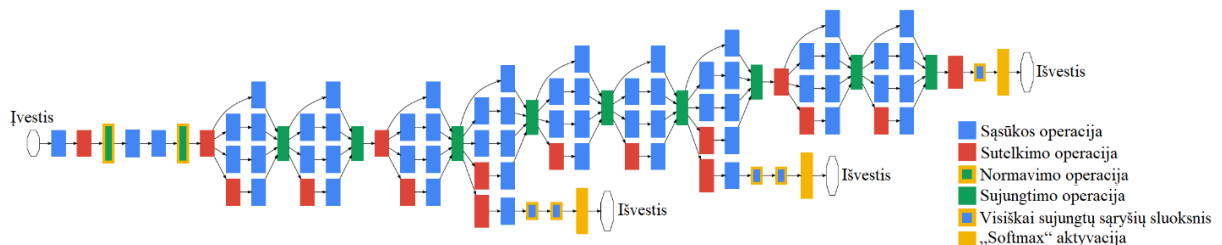
nykstančiojo gradiento problemą, o pats tinklas tapo vienu iš labiausiai kompiuterinės regos srityje naudojamų sąsūkos neuroninių tinklų [37, 38]. Be praleidimo jungčių „ResNet“ struktūra tampa artima „VGG“ tinklo struktūrai. Yra empirinių įrodymų, kad tinklą, su liekamaisiais ryšiais yra lengviau optimizuoti. Be to, gerokai padidinus gylį, galima gauti didesnę tikslumą [37]. 2015 m. „ResNet“ modelis, turintis 152 paslėptuosius sluoksnius, užėmė pirmąją vietą „ILSVRC2015“ konkurse. Modelio architektūra pateikta 1.5.2 paveiksle.



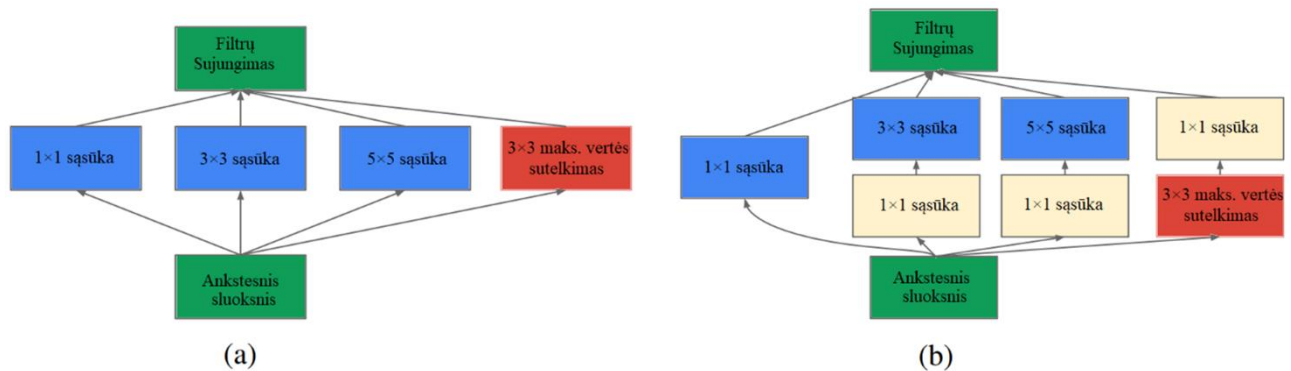
1.5.2 pav. „ResNet“ neuroninio tinklo architektūra [37]

1.5.3. „GoogLeNet“

Reikšminiai vaizdo požymiai gali būti įvairių dydžių ir labai skirtingi, todėl atliekant vaizdų klasifikavimo užduotį, parinkti tam tikro fiksuoto dydžio sąsūkos filtrą yra gana sudėtinga. Didesni filtrai labiau tinka globaliesiems požymiams, kurie yra pasiskirstę dideliame vaizdo plote. Kita vertus, mažesni filtrai užtikrina gerus rezultatus, aptinkant lokalias vaizdo savybes. Norint efektyviai atpažinti tokią kintamo dydžio požymį, reikia skirtingų dydžių filtrų. „GoogLeNet“ architektūra yra skirtingų idėjų derinys. Tinklą sudaro simetriniai ir asimetriniai blokai, įskaitant sąsūkos, sutelkimo ir VSS sluoksnius. Esminis elementas šioje architektūroje yra vadinamasis „Inception“ blokas (žr. 1.5.4. pav.), kuriame realizuojami keli skirtingų dydžių sąsūkos filtrai [39, 40]. Taigi, tinklas ne tik gilėja, bet ir plėtėja (žr. 1.5.3 pav.). „GoogLeNet“ buvo sukurtas 2014 m. ir nugalėjo „ILSVRC2014“ konkursą pasiekdamas žmogaus lygio vaizdų atpažinimo gebėjimus, naudojant 4 mln. parametrų.



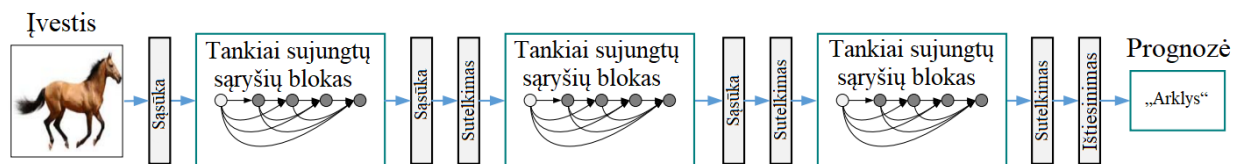
1.5.3 pav. „GoogLeNet“ neuroninio tinklo architektūra [39]



1.5.4 pav. „Inception“ blokas [39]

1.5.4. „DenseNet“

„Densenet“ architektūros pagrindinis elementas yra tankiai sujungtų sąryšių blokas. Šiame bloke kiekvienas sluoksnis gauna įvestį iš ankstesnių sluoksnių išvesties požymių žemėlapių. Kitaip nei „ResNet“ neuroniniame tinkle, čia požymių žemėlapiai yra sujungiami (angl. *Concatenated*), o ne sumuojami. Pakartotinis gautų požymių naudojimas pagerina gradientų sklidimą tinkle ir leidžia sumažinti sąsūkos filtrų skaičių, todėl šio tipo tinklai greičiau mokosi [41]. Sluoksnio išvesties požymių žemėlapių skaičius apibrėžiamas kaip augimo greitis (angl. *Growth Rate*), kuris kontroliuoja, kiek naujos informacijos kiekvienas sluoksnis įneša į tinklą. Supaprastinta tinklo architektūra pavaizduota 1.5.5 paveiksle.



1.5.5 pav. „DenseNet“ neuroninio tinklo architektūra [42]

1.6. Trumpas skyriaus apibendrinimas

Šiame skyriuje buvo susipažinta su dirbtinių neuronų tinklais ir giliaisiais neuroniniais tinklais. Aptarti klasikiniai vaizdų segmentavimo metodai. Išnagrinėti esminiai skirtingų architektūrų SNT, kurie istoriškai užėmė aukščiausias vietas vaizdų klasifikavimo ir objektų atpažinimo konkurse „ILSVRC“. Darytina išvada, kad DNT gerokai išplėtė segmentavimo galimybes. Pirma, DNT geba automatiškai išskirti reikšminius aukšto lygio požymius, todėl kompiuterinės regos inžinieriams nebereikia rašyti sudėtingų požymių išskyrimo algoritmų. Antra, DNT paremti segmentavimo metodai yra kur kas tikslesni nei klasikiniai. Trečia, DNT yra lankstesni, nes gali būti apmokomi iš naujo, naudojant kitus duomenis ir sprendžiant kitas problemas. Nagrinėtų SNT architektūrų pagrindu yra sukurta daug įvairių šiuolaikiškų sąsūkos neuroninių tinklų. Pažymima, kad vaizdų segmentavimo srityje šie tinklai yra vadinami enkoderiais (angl. *Encoders*) arba pamatiniais tinklais (angl. *Backbones*).

2. ATRINKTI GILIOJO MOKYMOŠI METODAI IR JŲ TAIKYMAS AUTOMATINEI AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZEI

Šiame skyriuje aptariami aerofotografinių nuotraukų segmentavimo iššūkiai ir modernūs giliojo mokymosi metodai, taikytini kompiuterinės regos srityje vaizdams segmentuoti. Taip pat pateikiamos įvairios tikslo funkcijos bei segmentavimo rezultatų įvertinimo metrikos. Skyriaus pabaigoje tiriama kokie šiuolaikiniai segmentavimo metodai geriausiai geba segmentuoti vienai ar kelioms klasėms priklausančius objektus aerofotografinėse nuotraukose. Nagrinėjami trys svarbūs aerofotografinių nuotraukų analizės uždaviniai – automatinis kelių, pastatų ir daugelio klasių segmentavimas, naudojant giliojo mokymosi metodus.

Apskritai, vaizdų segmentavimo problemą galima skirstyti į dvi dalis – dvejetainį ir daugelio klasių segmentavimą. Dvejetainio segmentavimo uždutis – išskirti vienos klasės objektus iš vaizdo fono, o daugelio klasių segmentavimo uždutis – išskirti skirtingoms klasėms priklausančius objektus iš vaizdo fono. Šiuolaikiniai pikselių lygio segmentavimo metodai dažniausiai naudoja giliųjų neuroninių tinklų architektūras, sudarytas iš dviejų elementų – enkoderio ir dekoderio. Enkoderis yra sąsūkos neuroninis tinklas, skirtas požymių žemėlapiams sudaryti, o dekoderis – tinklas, skirtas erdvinei semantinei informacijai atkurti [43]. Enkoderis paprastai būna inicijuojamas su filtrų svoriais, kurie yra gauti tinklą apmokius su labai dideliu duomenų rinkiniu, pavyzdžiui, „ImageNet“ [43]. Tokiu atveju yra „užšaldomi“ žemojo lygio sluoksniai, o mokomi tik keli paskutiniai enkoderio sluoksniai, kurie išskiria aukšto lygio požymius. Taip galima sutaupyti laiko ir skaičiavimo išteklių, o galutiniai rezultatai paprastai nežymiai skiriasi nuo tinklų, kurie buvo apmokyti nuo nulio, rezultatų [43].

2.1. Aeronuotraukų segmentavimo iššūkiai

Nuotolinio aptikimo vaizduose yra daug įvairių antžeminių objektų, tokių kaip keliai, pastatai, medžiai, transporto priemonės ir t. t. Dėl objektų įvairumo ir sudėtingo jų erdvinio pasiskirstymo, viso vaizdo segmentacija tampa sudėtingesnė problema. Iki šiol atlikti išsamūs nuotolinio aptikimo vaizdų segmentavimo tyrimai rodo, kad dar nėra tokio algoritmo, kuris galėtų tenkinamu tikslumu segmentuoti aerofotografinės nuotraukas [44]. Pagrindiniai keturi nuotolinio aptikimo vaizdų segmentavimo iššūkiai yra:

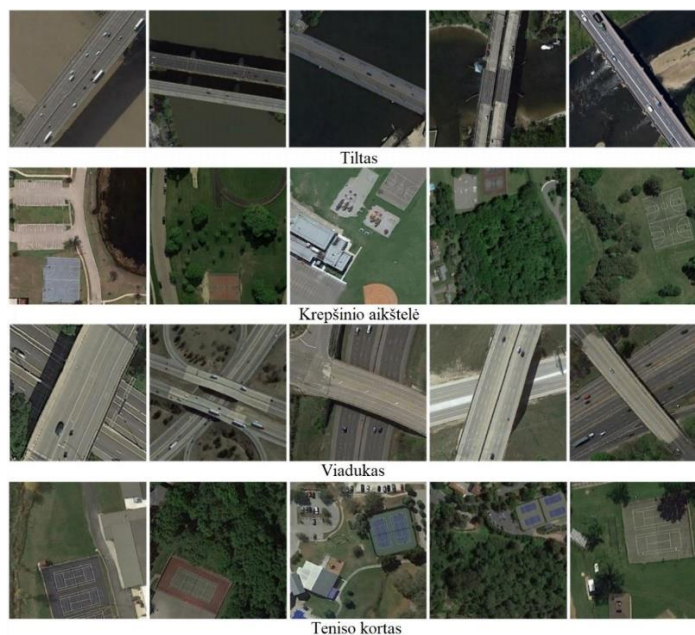
- didelė tos pačios klasės objektų įvairovė,
- didelis skirtingų klasių objektų panašumas,
- didelis objekto ir scenos mastelių skirtumas,
- kelių antžeminių objektų sambūvis.

Pirmasis iššūkis kyla dėl didelių žemės objektų pasirodymo toje pačioje semantinėje klasėje skirtumų (žr. 2.1.1 pav.). Antžeminiai objektai paprastai skiriasi stiliumi, forma, dydžiu ir erdvinio pasiskirstymu. Pavyzdžiui, bažnyčioms būdingi skirtingi stiliai, o oro uostai ir geležinkelio stotys išsiskiria savo formomis. Be to, kai palydovinės sistemos fiksuoja aerofotografines nuotraukas, toje pačioje semantinėje klasėje gali atsirasti didelių spalvų ir spinduliuotės intensyvumo skirtumų, dėl vaizdo gavimo sąlygų, kurioms įtakos gali turėti tokie veiksniai kaip oras, debesys, rūkas ir t. t. [44].



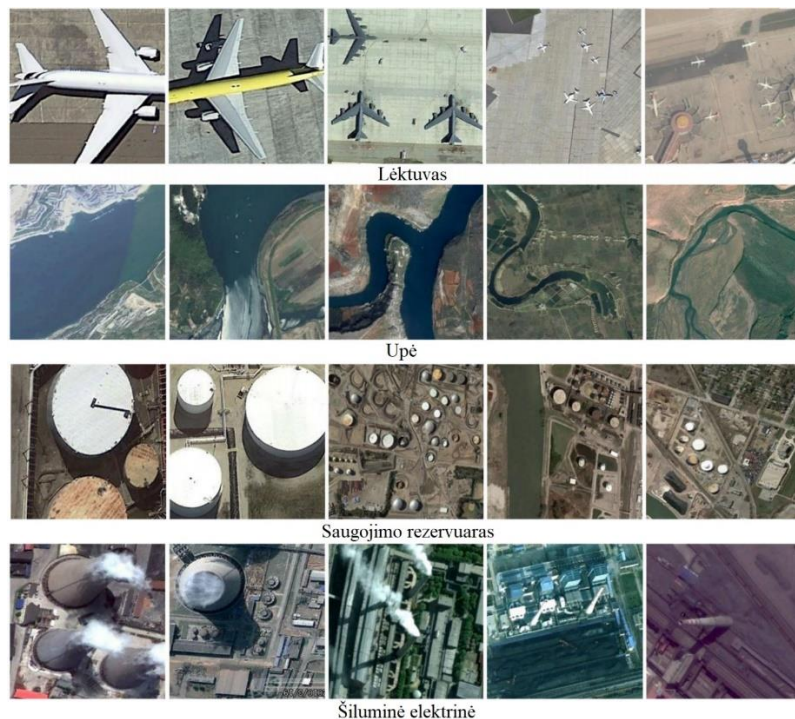
2.1.1 pav. Didelė tos pačios klasės objektų įvairovė [44]

Skirtingų klasių panašumo iššūkį dažniausiai sukelia objektų semantinės informacijos sutapimas. Pavyzdžiui, tiltai ir viadukai yra labai panašūs, bet priklauso skirtingoms klasėms. Krepšinio aikštelių ir teniso kortų semantinė informacija taip pat yra labai panaši [44]. Taigi, kai kurios sudėtingos vaizdo scenos yra panašios viena su kita savo vizualiniu turiniu, todėl klasių atskyrimas gali būti labai sudėtingas (žr. 2.1.2 pav.).



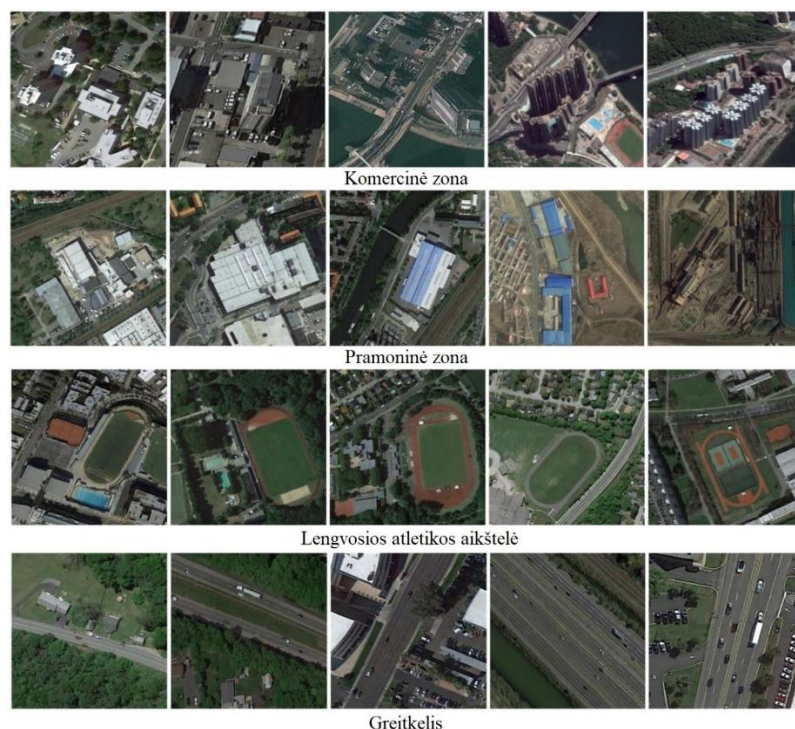
2.1.2 pav. Didelis skirtingų klasių objektų panašumas [44]

Objektų mastelių skirtumas taip pat yra svarbus nuotolinio aptikimo vaizdo segmentavimo iššūkis. Nuotolinio aptikimo jutikliai veikia įvairių aukščių orbitose – nuo kelių šimtų kilometrų iki daugiau nei dešimties tūkstančių kilometrų, tai lemia vaizdo aukščio kitimą. Be to, mastelių skirtumai gali atsirasti, dėl jutiklių vidinių parametų (angl. *Intrinsic Parameters*) skirtumų [44].



2.1.3 pav. Didelis objekto ir scenos mastelių skirtumas [44]

Paskutinė išvardinta problema atsiranda, dėl sudėtingo ir įvairaus antžeminių objektų pasiskirstymo ir plačios vaizdų perspektyvos, tad gana dažnai viename nuotolinio aptikimo vaizde yra keli antžeminiai objektai (žr. 2.1.4) [44].



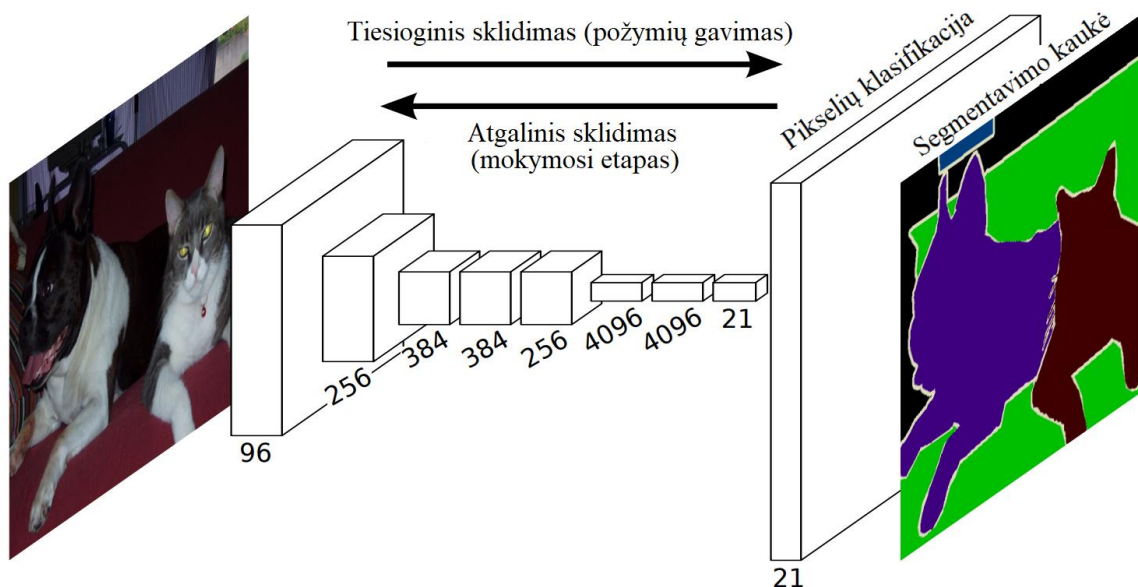
2.1.4 pav. Kelių antžeminių objektų sambūvis [44]

2.2. Segmentavimo modeliai

Šiame skirsnyje, remiantis [45, 46, 47, 48, 49, 50, 51], trumpai aprašomi populiarūs VKT, „U-Net“, FPN, „PSPNet“, „LinkNet“ ir „DeepLabV3+“ segmentavimo modeliai. Kiekvienas modelis iš esmės buvo sukurtas specifinėms segmentavimo problemoms spręsti, tačiau atlikus literatūros apžvalgą pastebėta, kad jie pritaikomi aerofotografinėms nuotraukoms segmentuoti.

2.2.1. VKT

Visiškai konvoliucinis tinklas yra tokia dirbtinių neuronų tinklų architektūra, kurioje naudojami tik sąsūkos, sutelkimo ir didinimo (angl. *Upsampling*) sluoksniai, atsisakant visiškai sujungtųjų sąryšių sluoksnio, todėl jie turi mažiau parametrų ir geba greičiau mokytis [47]. Be to, visiškai konvoliuciniame tinkle įvestys gali būti kintamo dydžio. Tiesioginio sklido metu yra gaunami požymiai, o klasifikavimo metu naudojamas didinimo sluoksnis, kuris leidžia atstatyti semantinę erdvinę informaciją. Siaurėjimo dalyje (angl. *Downsampling Path*) kiekvieno sluoksnio išvestyje yra prarandama dalis erdvinės informacijos, todėl yra panaudojamos praleidimo jungtys (angl. *Skip Connections*). 2.1.1 paveiksle pavaizduota VKT architektūra.

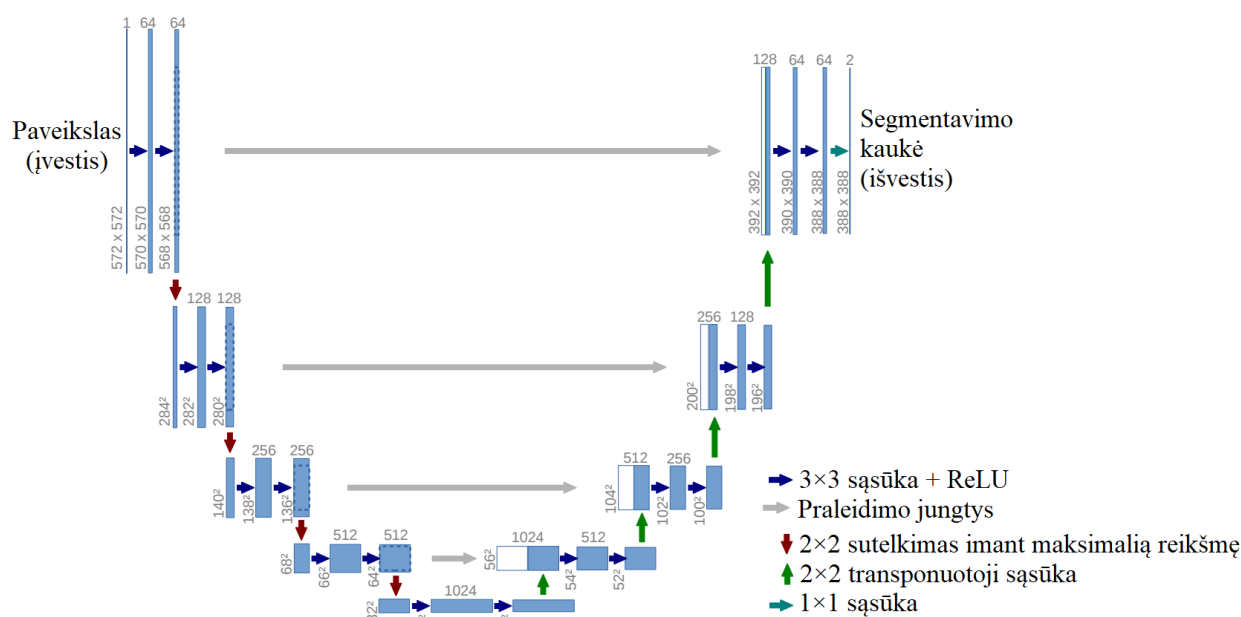


2.2.1 pav. VKT architektūra [47]

2.2.2. „U-Net“

„U-Net“ yra modifikuotas VKT modelis, kuris buvo sukurtas biomedicininiais vaizdams segmentuoti, tačiau šiandien jis tampa vis aktualesnis daugelyje kitų sričių, įskaitant nuotolinį aptikimą. Vėl gi, enkoderis naudojamas požymių gavimui, o dekoderis, atsižvelgdamas į lokalizaciją, sujungia išmokus požymius su erdvine semantine informacija [48]. Siaurėjimo dalyje yra pakartotinai naudojami du 3×3 dydžio sąsūkos filtrai, kuriems pritaikoma LF aktyvacijos funkcija ir 2×2 dydžio maksimalios vertės sutelkimo filtrai, su žingsniu (angl. *Stride*) lygiu 2. Kiekviename

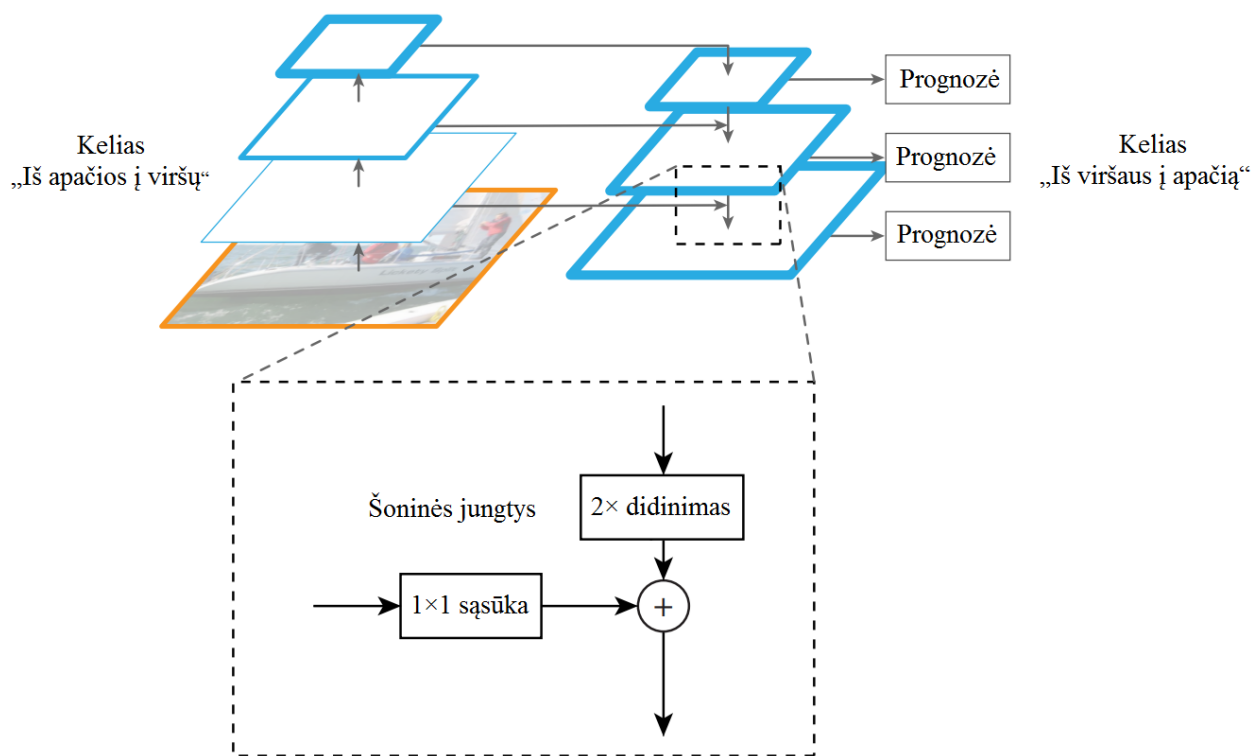
siaurėjimo etape yra dvigubinamas požymių žemėlapių (kanalų) skaičius. Platėjimo dalyje pakartotinai naudojami požymių žemėlapių didinimo sluoksniai, 2×2 dydžio transponuotosios sąsūkos filtrai, sujungimo filtrai (angl. *Concatenation*) ir 3×3 dydžio sąsūkos filtrai su LF aktyvacija. Transponuotosios sąsūkos filtrai perpus sumažina kanalų skaičių. Sujungimo sluoksnyje, naudojant praleidimo jungtis, esami požymių žemėlapiai sujungiami su atitinkamais apkarpytais požymių žemėlapiais, išskirtais siaurėjimo dalyje. Apkarpymas būtinas, dėl kraštinių pikselių praradimo po kiekvienos sąsūkos operacijos. Paskutiniame sluoksnyje naudojami 1×1 dydžio (1D) sąsūkos filtrai, kurie susieja požymius žemėlapių skaičių su klasių skaičiumi. Iš viso „U-Net“ tinklas turi 23 sluoksnius [48]. „U-Net“ architektūra pateikta 2.2.2 paveiksle.



2.2.2 pav. „U-Net“ architektūra [48]

2.2.3. FPN

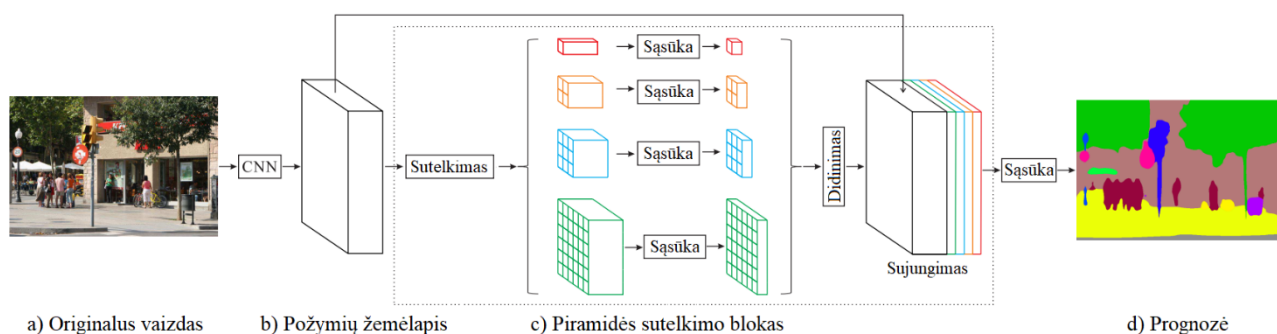
Požymių piramidės tinklas priima tam tikro fiksuoto dydžio vaizdą ir keliuose lygmenyse sudaro proporcingo dydžio požymių žemėlapius. Šis procesas yra nepriklausomas nuo enkoderių architektūrų. Požymių piramidės tinklą sudaro du keliai: „iš apačios į viršų“ ir „iš viršaus į apačią“. „Iš apačios į viršų“ kelias yra ne kas kita, kaip tiesioginio sklaidimo neuroninis tinklas, sudarantis požymių hierarchiją (skirtingo mastelio požymių žemėlapius nustatytu žingsniu). Kiekvieno etapo paskutinio sluoksnio išvestys naudojamos kaip kontrolinis požymių žemėlapių rinkinys. „Iš viršaus į apačią“ kelias sudaro aukštesnės raiškos požymius, didindamas abstraktesnius, aukštesnio piramidės lygio požymių žemėlapius, kurie, naudojant šonines jungtis (angl. *Lateral Connections*), yra sujungiami su to paties erdvinio dydžio požymių žemėlapiais, gautais kelyje „iš apačios į viršų“ (žr. 2.2.3 pav.). Kanalamis mažinti naudojami 1×1 dydžio sąsūkos filtrai. Trumpai tariant, FPN paima aukšto lygio vaizdo požymius iš pirmosios piramidės ir perduoda jas žemyn į apatinius sluoksnius. Tai leidžia susieti žemo ir aukšto lygio semantinę informaciją [49].



2.2.3 pav. FPN architektūra [49]

2.2.4. „PSPNet“

„PSPNet“ yra vaizdo segmentavimo modelis, kuris naudoja piramidės sutelkimo bloką globaliai kontekstinei informacijai gauti. Susiejus lokaliuosius ir globaliuosius požymius, galutinė prognozė tampa patikimesnė [50]. Pirmiausia, naudojant sąsūkos neuroninį tinklą, yra gaunamas požymių žemėlapis, kuris yra 8 kartus mažesnis už įvesties paveikslėlio dydį. Tuomet, Piramidės sutelkimo blokas, naudodamas 4 skirtingo dydžio sutelkimo ir 1×1 dydžio sąsūkos filtrus, išrenka įvairius vaizdo bruožus ir juos tarpusavyje sujungia. Taip gaunamas naujas požymių žemėlapis, turintis lokaliuosius ir globaliuosius vaizdo požymius. Pagaliau sąsūkos sluoksnis kiekvienam pikseliui sukuria galutinę prognozę (žr. 2.2.4 pav.).

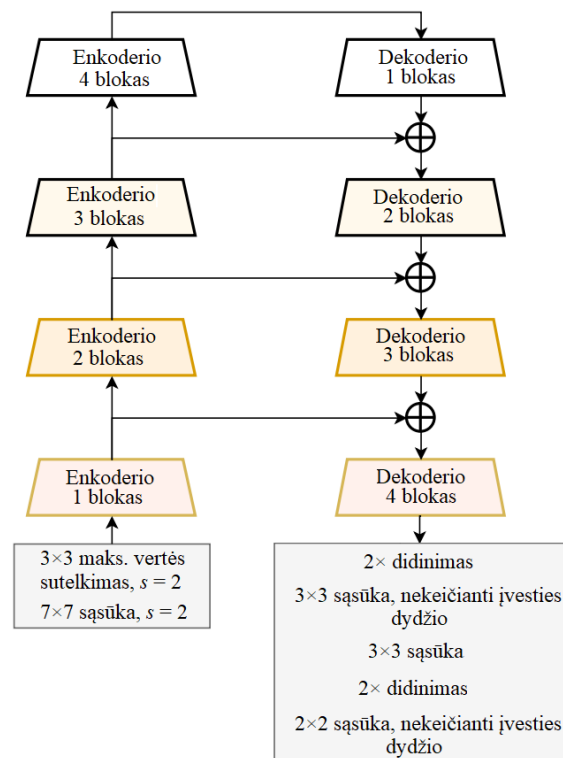


2.2.4 pav. „PSPNet“ architektūra [50]

2.2.5. „LinkNet“

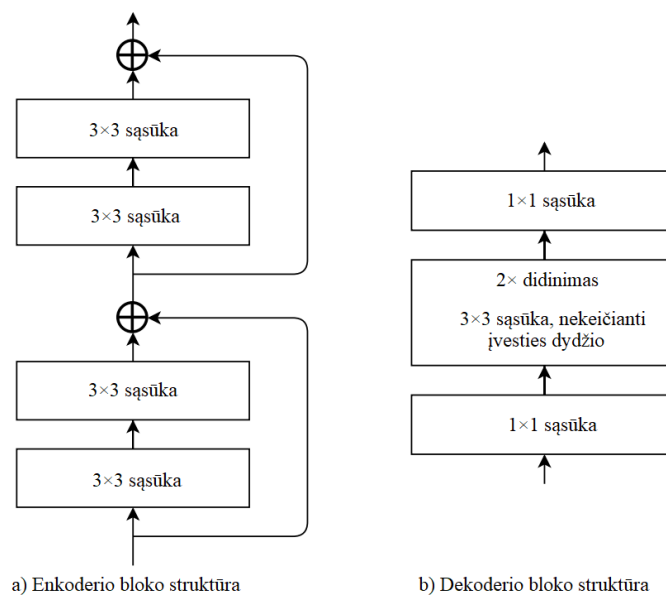
Šios architektūros neuroninis tinklas stengiasi efektyviai dalintis informaciją tarp enkoderio ir dekoderio bloką, naudodant papildomus ryšius. A. Chaurazijus ir kt. (Chaurasia et al., 2017) įrodė,

kad toks segmentavimo modelis ne tik rodo gerus rezultatus, bet yra spartesnis už kitus modelius, kadangi yra efektyviai naudojami parametrai.



2.2.5 pav. „LinkNet“ architektūra [51]

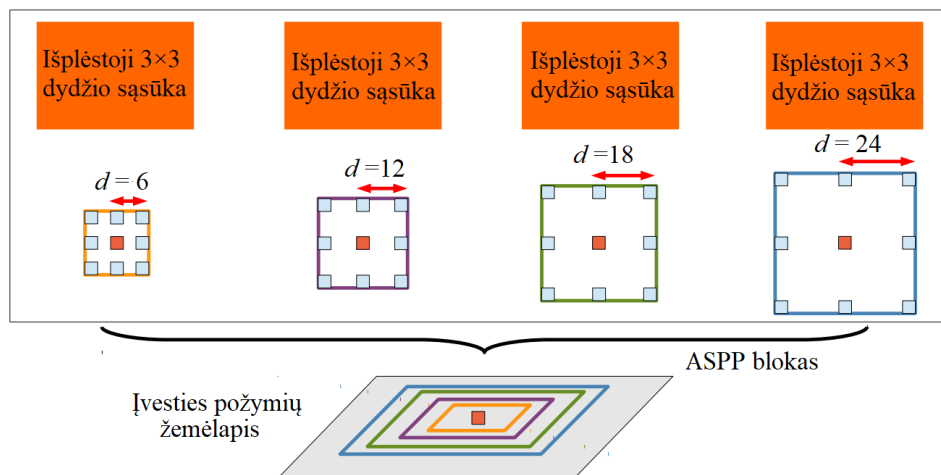
„LinkNet“ įvestyje yra atliekama 7×7 dydžio sąsūkos operacija, su žingsniu 2 ir 2×2 dydžio maksimalios vertės sutelkimo operacija, su žingsniu 2. Enkoderį sudaro 4 blokai, kuriuose atliekamos tik sąsūkos operacijos, o kiekviename bloke dvigubinamas filtrų skaičius. Iš esmės enkoderio struktūra atitinka „ResNet-18“ sąsūkos neuroninio tinklo struktūrą. Dekoderis taip pat sudarytas iš 4 blokų, kuriuose mažinamas filtrų skaičius, naudojant 1×1 dydžio sąsūkos operacijas. Išvesties bloke vaizdą klasifikuoja 2×2 dydžio sąsūkos filtras.



2.2.6 pav. Enkoderio ir dekoderio blokų struktūra [51]

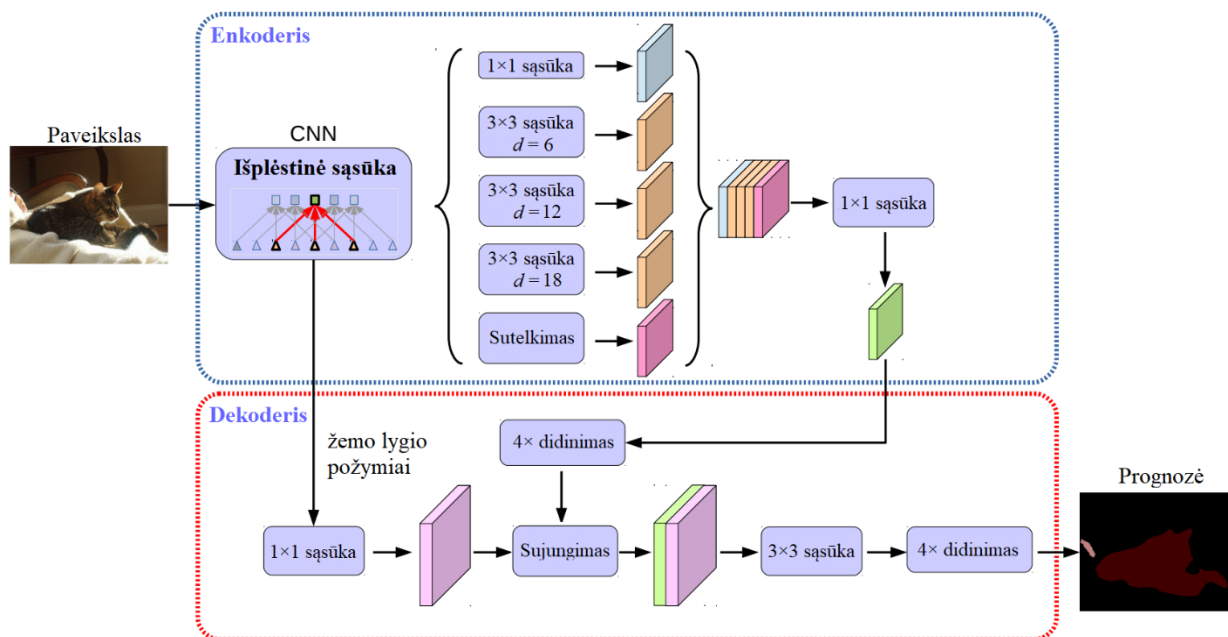
2.2.6. „DeepLabV3+“

„DeepLabV3+“ išsiskiria nuo kitų segmentavimo modelių tuo, kad yra vietoje klasikinių sąsukos filtrų naudojami išplėstinės sąsukos filtrai (angl. *Dilated Convolution*). Išplėstinės sąsukos operacija leidžia padidinti filtrų aprėptį, nekeičiant parametrų skaičiaus. Esminis „DeepLab“ modelio elementas – IEPS blokas (žr. 2.2.7 pav.).



2.2.7 pav. IEPS blokas [52]

IEPS blokas yra labai panašus į piramidės sutelkimo bloką, kuris naudojamas „PSPNet“ modelyje. Skirtumas tas, kad vietoje sutelkimo operacijos yra naudojama išplėstinės sąsukos operacija. IEPS bloką sudaro keli 3×3 dydžio sąsukos filtrai, su skirtingais išplėtimo žingsniais d . IEPS blokas padeda išskirti daugiaplanę kontekstinę informaciją. „DeepLabV3+“ segmentavimo modelio architektūra vaizduojama 2.2.8 paveiksle.



2.2.8 pav. „DeepLabV3+“ architektūra [53]

2.3. Tikslų funkcijos

Giliojo mokymosi algoritmai, optimizuodami klaidos funkciją, naudoja gradiento nusileidimo algoritmą. Norint tikslaus ir greito konvergavimo, reikia užtikrinti, kad matematiškai aprašomos tikslų funkcijos būtų vienareikšmiškai apibrėžtos ir galėtų apimti ekstremumus. Tikslų funkcijų atsiradimas siejamas su tradiciniu mašininio mokymusi, kai šios funkcijos buvo išvedamos pagal duomenų tikimybinis skirstinius. Pavyzdžiui dvejetainės kryžminės entropijos (angl. *Binary Cross Entropy*) tikslų funkcija yra kilusi iš Bernulio skirstinio. Apskritai, tikslų funkcijas galima suskirstyti į 4 kategorijas: paremtos pasiskirstymu (angl. *Distribution-based*), paremtos sritimis (angl. *Region-Based*) ir sudėtinės (angl. *Compounded*) (žr. 2.3.1 lentelę) [54].

2.3.1 lentelė. Tikslų funkcijų tipai [54]

Tipas	Tikslų funkcija
Paremtos pasiskirstymu	DKE Pasvertoji DKE Balansuotoji KE ŽTF
Paremtos sritimis	„Dice“ Jautrumo–specifiškumo Tverskio Židinio Tverskio
Sudėtinės	Kombinuotoji Eksponentinė–logaritminė

2.3.1. Dvejetainė kryžminė entropija

DKE funkcija apibrėžiama kaip dviejų tikimybių pasiskirstymo tam tikram atsitiktiniam kintamajam ar įvykių rinkiniui skirtumo matas. Ji plačiai naudojama klasifikavimo uždaviniui spręsti, o kadangi segmentacija yra pikselių lygio klasifikacija, tai ji tinka segmentavimo uždaviniui. Matematiškai DKE funkcija aprašoma taip:

$$L_{BCE}(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})), \quad (2.3.1)$$

čia y – tikroji vertė (angl. *Ground Truth*), o $\hat{y}$ – modelio prognozė.

2.3.2. Pasvertoji dvejetainė kryžminė entropija

Tai DKE funkcijos atmaina. Šiuo atveju teigiami pavyzdžiai įvertinami pagal tam tikrą nustatytą koeficientą. Ši funkcija plačiai taikoma iškreiptų ar nesubalansuotų duomenų atveju. Pasvertąją DKE galima apibrėžti taip:

$$L_{W-BCE}(y, \hat{y}) = -(\beta \cdot y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})), \quad (2.3.2)$$

čia buvo pridėtas, β – svorio koeficientas, kurį keičiant yra derinamas klaidingai nustatytų neigiamų elementų (KKNP) arba klaidingai nustatytų teigiamų elementų (KKTP) skaičius. Jei norima sumažinti klaidingai nustatytų neigiamų elementų skaičių, tai $\beta > 1$. Jeigu norima sumažinti klaidingai nustatytų teigiamų elementų skaičių, tai $\beta < 1$.

2.3.3. Balansuotoji kryžminė entropija

Šiuo atveju svorio koeficientas taikomas ne tik teigiamiems pavyzdžiams, bet ir neigiamiems, todėl matematinę išraišką galima aprašyti taip:

$$L_{BCE}(y, \hat{y}) = -(\beta \cdot y \log(\hat{y}) + (1 - \beta) \cdot (1 - y) \log(1 - \hat{y})). \quad (2.3.3)$$

Pažymima, kad $\beta = 1 - \frac{y}{H \cdot W}$, kur H ir W yra požymių erdvės dydis.

2.3.4. Židinio tikslo funkcija

Šią tikslo funkciją taip pat galima laikyti DKE funkcijos atmaina. ŽTF sumažina paprastai atskiriamų pavyzdžių indėlį ir leidžia modeliui daugiau dėmesio skirti sudėtingesnių pavyzdžių mokymuisi, todėl puikiai tinka labai nesubalansuotoms klasėms. Matematiškai ši funkcija yra išvedama iš DKE funkcijos taip:

$$CE = \begin{cases} -\log(p), & \text{kai } y = 1 \\ -\log(1 - p), & \text{kitais atvejais} \end{cases}, \quad (2.3.4)$$

ŽTF apskaičiuotą klasės tikimybę apibrėžia taip:

$$CE = \begin{cases} p, & \text{kai } y = 1 \\ 1 - p, & \text{kitais atvejais} \end{cases}, \quad (2.3.5)$$

tada KE galima perrašyti taip:

$$CE(p, y) = CE(p_t) = -\log(p_t). \quad (2.3.6)$$

Židinio tikslo funkcija savo ruožtu naudoja papildomą dėmenį – moduliacijos faktorių $a_t(1 - p_t)^\gamma$, todėl pilnoji formulė atrodo taip:

$$FL(p_t) = -a_t(1 - p_t)^\gamma \log(p_t), \quad (2.3.7)$$

čia koeficientas $\gamma > 0$, o kai $\gamma = 1$, ŽTF veikia lygiai kaip kryžminės entropijos funkcija. Taip pat pažymima, kad α kinta ribose $[0, 1]$.

2.3.5. „Dice“ tikslo funkcija

„Dice“ koeficientas plačiai naudojamas kompiuterinės regos bendruomenėje, nustatant dviejų vaizdų panašumą. Šis koeficientas buvo pritaikytas „Dice“ tikslo funkcijai sudaryti.

$$DL(y, \hat{p}) = 1 - \frac{2y\hat{p} + 1}{y + \hat{p} + 1}, \quad (2.3.8)$$

čia trupmenos skaitiklyje ir vardiklyje yra pridėti vienetai, kad funkcija būtų apibrėžta kraštutiniais atvejais, pvz., $y = \hat{p} = 0$.

2.3.6. Tverskio tikslo funkcija

Tverskio indeksas gali būti vertinamas kaip „Dice“ koeficiento apibendrintas atvejis. Naudojant β koeficientą, klaidingai nustatytiems teigiamiems ir neigiamiems elementams yra pridedamas svoris.

$$TI(p, \hat{p}) = \frac{p\hat{p}}{p\hat{p} + \beta(1-p)\hat{p} + (1-\beta)p(1-\hat{p})} \quad (2.3.9)$$

Pavyzdžiui, kai $\beta = 1/2$, Tverskio indeksas tampa įprastu „Dice“ koeficientu. Tverskio tikslo funkciją galima aprašyti taip:

$$TL(p, \hat{p}) = 1 - \frac{1 + p\hat{p}}{1 + p\hat{p} + \beta(1-p)\hat{p} + (1-\beta)p(1-\hat{p})} \quad (2.3.10)$$

2.3.7. Tverskio židinio tikslo funkcija

Panašiai kaip ir (2.1.7), šia tikslo funkcija bandoma didesnę dėmesį skirti sudėtingesniems pavyzdžiams, tačiau tik tam tikrose dominančiose srityse.

$$FTL = \sum_c (1 - TI_c)^\gamma, \quad (2.3.11)$$

čia TI – Tverskio indeksas, o koeficientas γ kinta ribose nuo 1 iki 3.

2.3.8. Jautrumo ir specifiškumo tikslo funkcija

Jautrumas (angl. *Sensitivity*) ir specifiškumas (angl. *Specificity*) yra plačiai naudojami vertinimo kriterijai. Sudarant aprašomąją tikslo funkciją, naudojamas parametras W , kuris sprendžia klasių disbalanso problemą. Pati funkcija apibrėžiama taip:

$$SSL = \omega \cdot \text{jautrumas} + (1 - \omega) \cdot \text{specifiškumas}, \quad (2.3.12)$$

$$\text{jautrumas} = \frac{TP}{TP + FN}, \quad (2.3.13)$$

$$\text{specifiškumas} = \frac{TN}{TN + FP}, \quad (2.3.14)$$

čia $TKTP$ – teisingai klasifikuoti teigiami pavyzdžiai, $KKNP$ – klaidingai klasifikuoti neigiami pavyzdžiai, $TKNP$ – teisingai klasifikuoti neigiami pavyzdžiai, $KKTP$ – klaidingai klasifikuoti teigiami pavyzdžiai.

2.3.9. Jungtinė tikslo funkcija

Jungtinė (kombinuotoji) tikslo funkcija yra svertinės „Dice“ ir modifikuotos KE funkcijų kombinacija. Šia funkcija bandoma padidinti lankstumą, sprendžiant klasių disbalanso problemą, kartu glotninant funkcijos kreivę. Jungtinė tikslo funkcija apibrėžiama taip:

$$L_{m-bce} = -\frac{1}{N} \sum_i \beta(y - \log(\hat{y})) + (1 - \beta)(1 - y) \log(1 - \hat{y}), \quad (2.3.15)$$

$$CL(y, \hat{y}) = \alpha L_{m-bce} - (1 - \alpha) DL(y, \hat{y}), \quad (2.3.16)$$

čia DL – „Dice“ funkcija.

2.3.10. Eksponentinė–logaritminė tikslo funkcija

Ši tikslo funkcija orientuota į ne taip tiksliai nuspėjamas vaizdo struktūras, naudojant mišrią „Dice“ ir KE formulę. Buvo pasiūlyta taikyti eksponentines ir logaritmines transformacijas minėtoms tikslo funkcijoms, kad būtų atsižvelgta į smulkesnių sprendimų ribas ir tikslų duomenų paskirstymą. Matematiškai ši funkcija aprašoma taip:

$$L_{Exp} = \omega_{Dice} L_{Dice} + \omega_{cross} L_{cross}, \quad (2.3.17)$$

$$L_{Dice} = E(-\ln(DC))^{\gamma_{Dice}}, \quad (2.3.18)$$

$$L_{cross} = E(\omega_l(-\ln(p_l))^{\gamma_{cross}}). \quad (2.3.19)$$

Apibendrinant nagrinėtas funkcijas, buvo sudaryta lentelė, kurioje nurodyti funkcijų taikymo atvejai.

2.3.2 lentelė. Įvairiems vaizdams segmentuoti naudojamų tikslo funkcijų apibendrinimas [54]

Tikslo funkcija	Taikymas
DKE	Geriausiai veikia kai klasės yra vienodai pasiskirsčiusios – balansuotos
Pasvertoji KE	Taikoma, kai yra klasių disbalansas. Svorio koeficientas taikomas teigiamiems pavyzdžiams
Balansuotoji KE	Taikoma, kai yra klasių disbalansas. Svorio koeficientas taikomas teigiamiems ir neigiamiems pavyzdžiams
Židinio	Geriausiai veikia su labai nesubalansuotu duomenų rinkiniu. Didesnis dėmesys skiriamas mokytis sudėtingesnius pavyzdžius
„Dice“	Taikoma, kai reikia įvertinti segmentavimo rezultatus
Jautrumo-specifiškumo	Naudojama tais atvejais, kai daugiau dėmesio skiriama teisingai klasifikuotiems teigiamiems elementams
Tverskio	Tai „Dice“ koeficiento atmaina, kurioje yra pridedamas svoris klaidingai nustatytiems teigiamiems bei neigiamiems elementams
Tverskio židinio	Tai Tverskio funkcijos modifikacija, kurioje didesnis dėmesys skiriamas sudėtingesniems pavyzdžiams
Kombinuotoji	Taikoma tada, kai yra nedidelis klasių disbalansas, siekiant išnaudoti „Dice“ ir DKE funkcijų privalumus
Eksponentinė-Logaritminė	Didesnis dėmesys skiriama ne taip tiksliai nuspėtomis klasėms

Tikslo funkcijos vaidina esminį vaidmenį, nustatant modelio efektyvumą. Tokiems sudėtingiems uždaviniams kaip vaizdų segmentavimas, nėra vienos universalios tikslo funkcijos. Dažniausiai tai priklauso nuo mokymo tikslais naudojamų duomenų rinkinio ypatybių, pvz., paskirstymo, kreivės asimetrijos, ribų ir t. t. Nei viena iš minėtų praradimo funkcijų nėra efektyviausia visais naudojimo atvejais. Tačiau galima sakyti, kad labai nesubalansuotiems duomenims geriau taikyti sutelktąsias tikslo funkcijas. Balansuotiems duomenims puikiai tinka ir KE funkcija, o šiek tiek nebalansuotiems duomenims galima taikyti apibendrintą „Dice“ koeficientą.

2.4. Įvertinimo metrikos

Idealiu atveju modelis turėtų būti vertinamas keliais kriterijais, pvz., kiekybiniu tikslumu, greičiu (sprendimo laiku) ir saugojimo reikalavimais. Tačiau dauguma iki šiol atliktų tyrimų orientuoti tik į modelio tikslumo vertinimo metriką. S. Minajė ir kt. (Minajė et al., 2020) apibendrina populiariausias metrikas, skirtas segmentavimo algoritmų tikslumui įvertinti. Nors kiekybinė metrika naudojama įvairiems lyginamųjų indeksų modeliams palyginti, sprendžiant, kuris modelis yra geriausias, svarbi modelio rezultatų vizualinė kokybė (kadangi žmogui vizuali informacija yra labiau suprantama).

2.4.1. Pikselių tikslumas (PA)

Pikselių tikslumo kriterijus paprasčiausiai parodo kiek pikselių buvo klasifikuota teisingai. Pavyzdžiui, turint $K + 1$ klasių (K objektų klasių ir fonas), matematiškai ši metrika aprašoma taip:

$$PA = \frac{\sum_{i=0}^K p_{ii}}{\sum_{i=0}^K \sum_{j=0}^K p_{ij}}, \quad (2.4.1)$$

čia p_{ij} yra pikselių, priklausančių i klasei, bet nustatytų kaip j klasė, skaičius. Vidutinis pikselių tikslumas yra išplėstinis PA variantas, kuriame teisingų pikselių santykis apskaičiuojamas kiekvienai klasei ir išvedamas vidurkis pagal bendrą klasių skaičių.

$$MPA = \frac{1}{K + 1} \sum_{i=0}^K \frac{p_{ii}}{\sum_{j=0}^K p_{ij}} \quad (2.4.2)$$

2.4.2. IoU

Šis kriterijus, dar žinomas kaip Džakardo (angl. *Jaccard*) indeksas, yra viena dažniausiai naudojamų semantinio segmentavimo metrikų. Ji apibrėžiama kaip žmogaus sužymėtų segmentų ir modelio spėtų segmentų persidengimo sritis, padalyta iš žmogaus sužymėtų segmentų ir modelio spėtų segmentų sąjungos srities:

$$IoU = J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (2.4.3)$$

Daugelio klasių segmentavimo atveju yra naudojamas vidutinis IoU (mIoU) įvertis.

2.4.3. Preciziškumas, jautrumas ir „F1“ įvertis

Šie kriterijai savotiškai įvertina modelio tikslumą, todėl įvertinimo lentelėje dažniausiai nurodomi kartu.

$$\text{Preciziškumas} = \frac{TP}{TP + FP}, \quad (2.4.4)$$

$$\text{Jautrumas} = \frac{TP}{TP + FN}, \quad (2.4.5)$$

$$F1 = \frac{2 \cdot \text{Preciziškumas} \cdot \text{Jautrumas}}{\text{Preciziškumas} + \text{Jautrumas}}, \quad (2.4.6)$$

čia *TKTP* nurodo teisingai klasifikuotų teigiamų elementų dalį, *KKTP* nurodo klaidingai klasifikuotų teigiamų elementų dalį, o *KKNP* – klaidingai klasifikuotų neigiamų dalį. „F1“ įvertis yra preciziškumo ir jautrumo metrikų harmoninis vidurkis.

2.4.4. „Dice“ koeficientas

Dar viena populiari vaizdų segmentavimo metrika (dažniau naudojama medicininių vaizdų analizėje), kurią galima apibrėžti kaip santykį tarp dvigubos žmogaus sužymėtų segmentų ir modelio spėtų segmentų persidengimo srities, ir bendro pikselių skaičiaus abiejuose vaizduose. Iš principo „Dice“ koeficientas panašus į IoU:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}. \quad (2.2.26)$$

Kai yra atliekama dvejetainė segmentacija (atskiriami vienos klasės objektai nuo fono) ir objektai priskiriami teigiamiems elementams, tuomet „Dice“ koeficientas tampa „F1“ įverčiu:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} = F1. \quad (2.2.27)$$

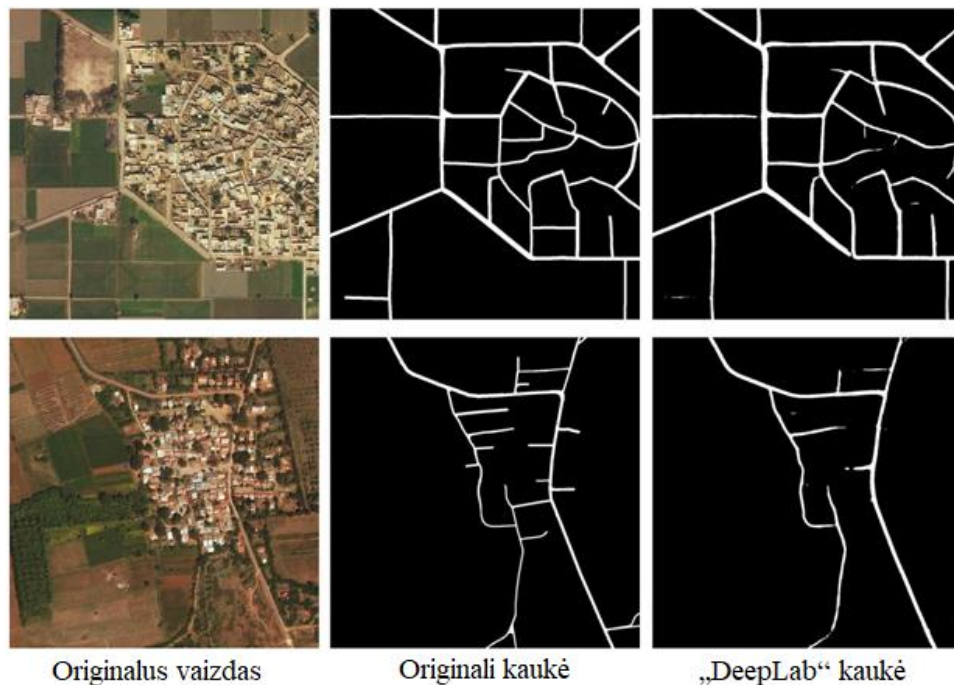
2.5. Segmentavimo modelių tyrimas dvejetainio segmentavimo atveju

Atliekant literatūros analizę pastebėta, kad dvejetainio segmentavimo atveju daugiausia tyrimų buvo atlikta, siekiant automatiškai išskirti kelius arba pastatus iš aerofotografinių nuotraukų. Šiame skirsnyje pateikiami ir nagrinėjami kompiuterinės regos ir nuotolinio aptikimo sričių bendruomenės tyrėjų gauti rezultatai. Išskiriami geriausius rezultatus rodantys segmentavimo modeliai. Akcentuojamos problemos, su kuriomis susidūrė tyrėjai.

2.5.1. Kelių segmentavimas

Keliai būna įvairių tipų (asfaltiniai, betoniniai, žvyro ir t. t.) ir formų (platūs, siauri, tiesūs, vingiuoti). Aeronuotaukų perspektyvoje kelius dažnai dengia tanki augmenija, pastatai, šešėliai ir kt. Šios priežastys apsunkina kelių segmentavimo uždavinį. [55].

I. Demir ir kt. (Demir et al., 2018) atliko kelių segmentavimo eksperimentus, naudodami savo sudarytą „DeepGlobe“ aerofotografinių nuotraukų rinkinį. Geriausi rezultatai buvo gauti, taikant modifikuotą „DeepLab“ modelį, turintį „ResNet18“ enkoderį. Tinklui apmokyti, autoriai naudojo židinio tikslo funkciją. Pabrėžiama, kad duomenims papildyti buvo naudojama tik vaizdo pasukimo transformacija, o tarpiniams rezultatams apdoroti pritaikomas dvejetainis slenkstis (vertės mažesnės už 128 priskiriamos fonui, o likusios – keliams). Segmentavimo rezultatams įvertinti buvo naudotas IoU rodiklis. Po 100 epochų mokymo, modelis pasiekė 54,5 % IoU įvertį. [5].



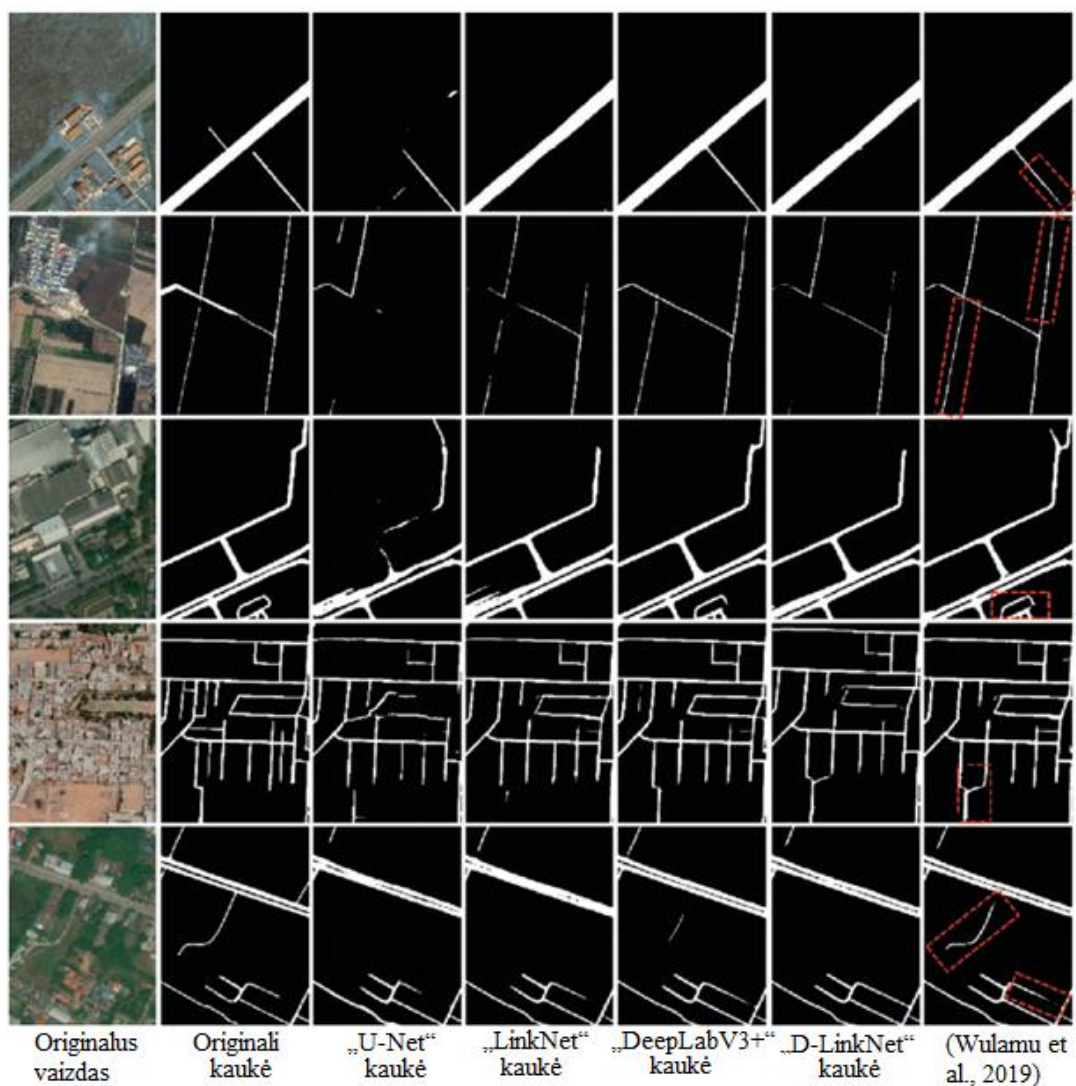
2.5.1 pav. Kelių segmentavimo, naudojant „DeepGlobe“ duomenų rinkinį, rezultatų vaizdavimas [5]
 Autoriai pastebėjo, kad modelis sunkiau išskiria trumpesnius ir siauresnius kelius (žr. 2.5.1 pav.). Be to, vietomis keliai nyksta, todėl tolimesniems tyrimams buvo pasiūlyta vietoje slenksčio taikyti kitus papildomo apdorojimo metodus.

A. Vulamu ir kt. (Wulamu et al., 2019), sprendami kelių segmentavimo uždavinį, pasiūlė naują neuroninio tinklo architektūrą. Autoriai modifikavo „U-Net“ modelį, pridėdami IEPS bloką. Yra žinoma, kad IEPS blokas padeda gauti globalią kontekstinę informaciją, todėl turėtų pagerinti aerofotografinių nuotraukų segmentavimo rezultatus. Eksperimentams atlikti autoriai naudojo Inčuanuo miesto kelių duomenų rinkinį ir pabrėžė, kad duomenys yra nesubalansuoti. Dėl pastarosios priežasties, remiantis [57], buvo sudaryta specifinė tikslo funkcija, sujungianti DKE, „Dice“ ir Lovaso funkcijas.

2.5.1 lentelė. Kelių segmentavimo rezultatai, naudojant Inčuanio kelių duomenų rinkinį [56]

Metodas	mIoU, %	PA, %
„U-Net“	79,3	97,8
„LinkNet“	81	98,1
„DeepLabV3+“	83,1	98,3
„D-LinkNet“	82,1	98,2
(Wulamu et al., 2019)	82,3	98,3

Mokymo etape, siekdami išvengti persimokymo, autoriai mokymo duomenų rinkiniui papildyti naudojo atsitiktinio pasukimo, horizontalaus ir vertikalios apvertimo, mastelio keitimo ir postūmio transformacijas. Rezultatams įvertinti buvo panaudotos mIoU ir PA metrikos. 2.5.1 lentelėje pateikiami skirtingų segmentavimo modelių rezultatai, o 2.5.2 paveiksle nurodomi vaizdiniai skirtingų modelių rezultatų skirtumai. Autoriai teigia, kad jų sukurtas metodas yra atsparesnis kelių trūkių problemoms, kurias sukelia šešėliai, okliuzijos ir pan. Pabrėžiama, kad originalios kaukės nėra visiškai tikslios, o tai lemia prastesnius rezultatus [56].

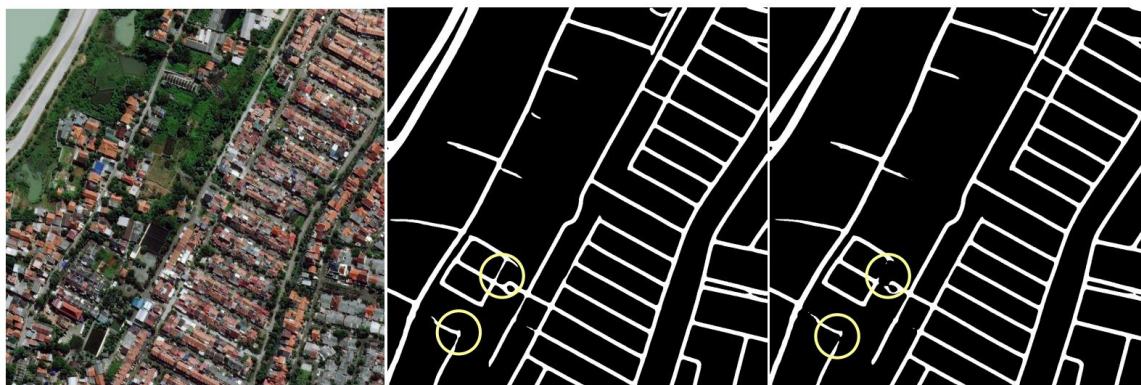
**2.5.2 pav.** Kelių segmentavimo, naudojant Inčuanio duomenų rinkinį, rezultatų vaizdavimas [56]

A. Tranas ir kt. (Tran et al., 2020), siekdami rasti kompromisą tarp modelio spartos ir tikslumo, pasiūlė naudoti modifikuotą „LinkNet“ modelį. Autoriai paskutiniame enkoderio sluoksnyje pridėjo piramidės sutelkimo bloką. Modifikuotas modelis buvo pavadintas „PP-LinkNet“. Keliams segmentuoti buvo pasirinktas „DeepGlobe“ duomenų rinkinys. Atlikdami papildomus eksperimentus, autoriai išskyrė židinio tikslo funkciją kaip vieną veiksmingiausių keliams segmentuoti. Pabrėžiama, kad mokymo metu buvo mažinamas mokymo žingsnis (angl. *Learning Rate*). Rezultatams įvertinti buvo taikyta mIoU metrika. Tirtų modelių segmentavimo rezultatai pateikti 2.5.2 lentelėje.

2.5.2 lentelė. Kelių segmentavimo rezultatai, naudojant „DeepGlobe“ duomenų rinkinį [58]

Metodas	Enkoderis	Tikslo funkcija	mIoU, %
„TernausNetV2“	-	ŽTF	47,16
„PSPNet“	-	ŽTF	63,38
„D-LinkNet“	„ResNet34“	„Dice“	66,31
		DKE	67,21
		ŽTF	67,82
„PP-LinkNet34“	„ResNet34“	ŽTF	70,53

Matyti, kad „PP-LinkNet“ viršijo kitų modernių segmentavimo modelių rezultatus, pasiekiant 70,53 % IoU įvertį. Norėdami pabrėžti piramidės sutelkimo bloko svarbą, autoriai pavaizdavo segmentavimo rezultatų skirtumus. (žr. 2.5.3 pav.). Taip pat buvo išvelgtos geros apibendrinimo galimybės, atlikus pastatų segmentavimą ir gavus artimus moderniems modeliams rezultatus [58].



2.5.3 pav. Segmentavimo rezultatų palyginimas naudojant piramidės sutelkimo bloką (viduryje) ir nenaudojant jo (dešinėje) [58]

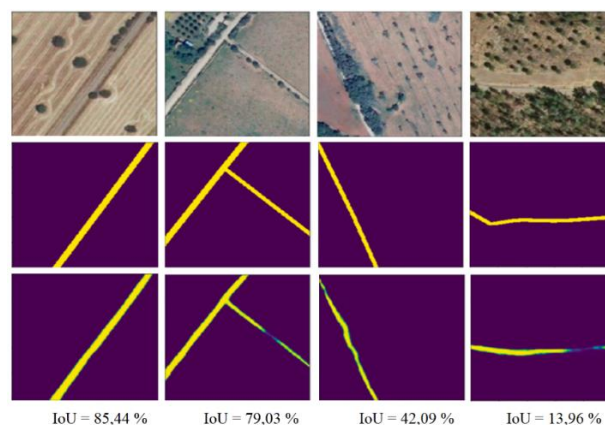
C. Cira ir kt. (Cira et al., 2020) atliko išsamius tyrimus, naudodami „U-Net“, FPN ir „LinkNet“ segmentavimo modelius. Kiekvienam modeliui buvo pritaikyti „MobileNetV2“, „EfficientNet“, „InceptionV3“, „VGG16“ ir „SEResNetXt50“ enkoderiai. Eksperimentams atlikti buvo pasirinktas „PNOA“ Ispanijos kelių duomenų rinkinys. Duomenims papildyti buvo naudojamos pasukimo, horizontalaus ir vertikalios apvertimo, postūmio, mastelio keitimo transformacijos. Be to, buvo keičiamas įvesties paveikslų ryškumas ir kontrastas. Autoriai naudojo modifikuotą tikslo funkciją

sujungdami DKE ir Džakardo funkcijas. Rezultatams įvertinti buvo taikomi vidutiniai IoU ir „F1“ įverčiai. 2.5.3 lentelėje pateikiami tirtų modelių kelių segmentavimo rezultatai.

2.5.3 lentelė. Kelių segmentavimo rezultatai, naudojant „PNOA“ duomenų rinkinį [55]

Metodas	Enkoderis	mIoU, %	vidutinis F1 įvertis, %
„U-Net“	-	52,99	66,38
	„MobileNetV2“	53,35	66,94
	„EfficientNet“	55,93	69,58
	„InceptionV3“	50,28	66,41
	„VGG16“	54,66	68,19
	„SEResNetXt50“	56,44	69,88
„LinkNet“	-	51,84	65,52
	„MobileNetV2“	53,02	66,83
	„EfficientNet“	56,26	69,79
	„InceptionV3“	52,93	66,63
	„VGG16“	53,72	67,44
	„SEResNetXt50“	55,15	68,77
FPN	-	52,33	66,21
	„MobileNetV2“	52,98	66,50
	„EfficientNet“	54,37	68,27
	„InceptionV3“	52,76	66,56
	„VGG16“	54,21	67,81
	„SEResNetXt50“	55,09	68,71

Matyti, kad naudojant „U-Net“ ir „SEResNetXt50“ konfigūraciją buvo pasiekti geriausi kelių segmentavimo rezultatai. 2.5.4 paveiksle parodoma, kaip IoU rodiklis įvertina segmentavimo kokybę. Pirmosios dvi eilutės rodo originalius vaizdus ir originalias kaukes, o trečioji eilutė rodo minėtos konfigūracijos sistemos prognozuotas kaukes.



2.5.4 pav. Kelių segmentavimo, naudojant „PNOA“ duomenų rinkinį, rezultatų vaizdavimas [55]

N. Abderahimas su bendraautoriais (Abderrahim et al., 2020) akcentuodami „U-Net“ modelio sėkmę medicinos srityje, atliko kelių segmentavimo tyrimus, kuriuose lygino VKT, „RSRCNN“, „SegNet“ ir „U-Net“ modelių galimybes. Eksperimentams atlikti autoriai naudojo Masačusetso kelių duomenų rinkinį. Modeliai buvo apmokomi taikant DKE tikslo funkciją. Duomenims papildyti buvo taikomos vaizdo pasukimo, apvertimo, mastelio keitimo transformacijos. Rezultatams įvertinti buvo pasitelktos „F1“, preciziškumo, specifiškumo ir PA metrikos. Tirtų modelių rezultatai pateikti 2.5.4 lentelėje.

2.5.4 lentelė. Kelių segmentavimo rezultatai, naudojant Masačusetso kelių duomenų rinkinį [59]

Metodas	„F1“, %	Preciziškumas, %	Specifiškumas, %	PA, %
VKT	53,2	43,5	68,6	90,4
„RSRCNN“	66,2	60,6	72,9	92,4
„SegNet“	76,8	77,3	76,5	95,7
„U-Net“	87,5	86,8	88,3	97,7

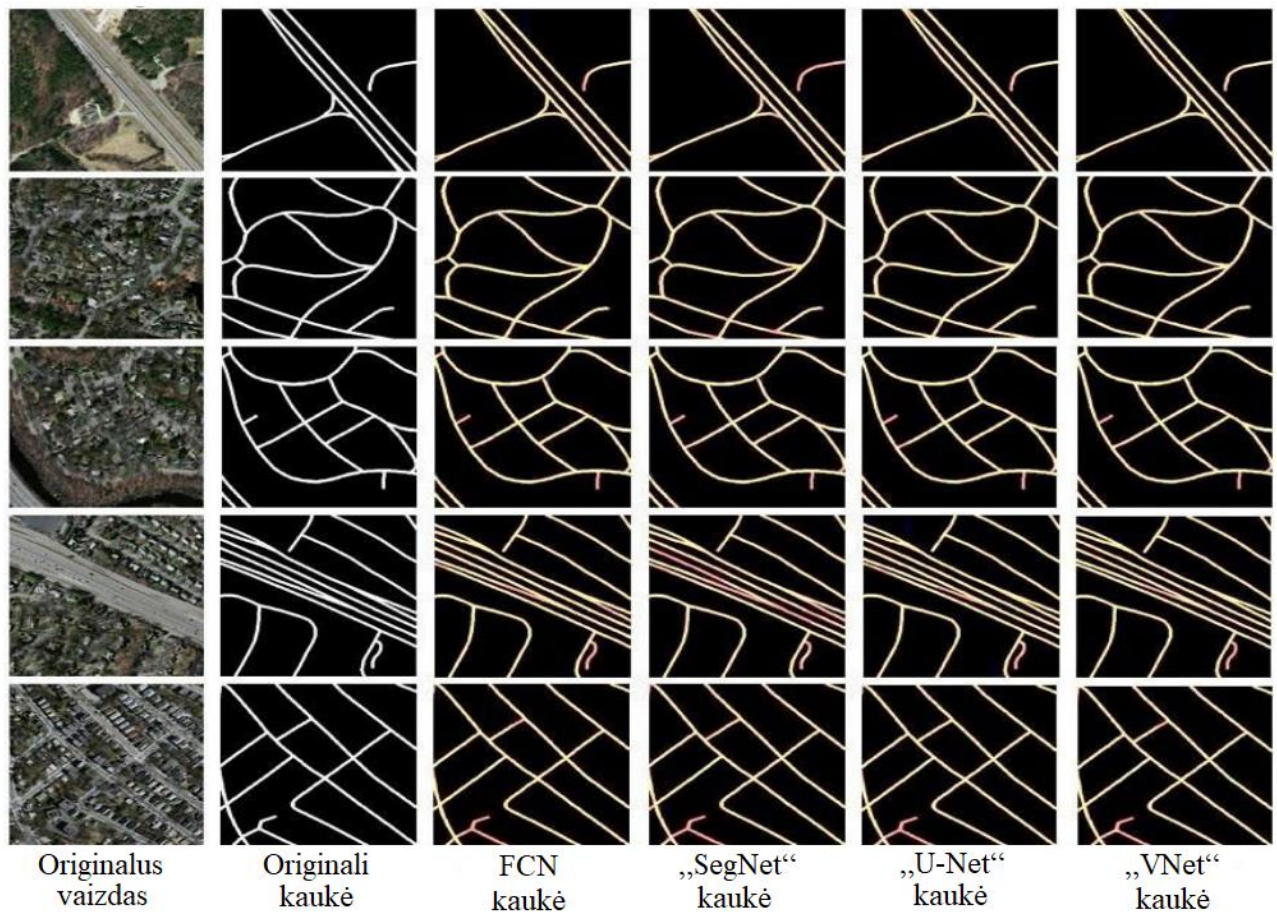
Matyti, kad „U-Net“ veiksmingumas pasitvirtino nagrinėjamu atveju. Lygindami rezultatus, autoriai pabrėžė, kad „U-Net“ tiksliau aptinka kiekvieną eismo juostą ir sudėtingas kelių struktūras. Modelis sugeba atskirti objektus, struktūriškai panašius į kelius, pavyzdžiui, automobilių stovėjimo aikštes ir pastatų stogus. Be to, gali segmentuoti kelius, kuriuos dengia medžiai.

A. Abdolahi ir kt. (Abdollahi et al., 2020) pristatė naują giliojo mokymosi modelį „VNet“, skirtą keliams segmentuoti. Eksperimentams atlikti buvo pasirinktas Masačusetso kelių duomenų rinkinys. Duomenų disbalanso problemai spręsti buvo pasitelkta jungtinė tikslo funkcija. Rezultatams įvertinti buvo naudojamos IoU ir „F1“ metrikos. „VNet“ kelių segmentavimo rezultatai buvo lyginami su VKT, „SegNet“, ir „U-Net“ modelių rezultatais. Tiesa, buvo pateikti 5 aerofotografinių nuotraukų segmentavimo rezultatai (žr. 2.5.5 lentelę). 2.5.5 paveiksle pateiktos lyginamųjų modelių prognozuotos kaukės. Raudona spalva paryškinti klaidingai segmentuoti vaizdo taškai.

2.5.5 lentelė. Kelių segmentavimo rezultatai, naudojant Masačusetso kelių duomenų rinkinį [60]

Metodas	IoU, %	„F1“, %
VKT	83,81	91,2
„SegNet“	80,43	89,16
„U-Net“	86,62	92,83
„VNet“	87,16	93,14

Kadangi autoriai pateikė tik 5 aerofotografinių nuotraukų segmentavimo rezultatus, sunku įvertinti bendrus rezultatus. Spėjama, kad vidutinė IoU vertė gerokai sumažėtų, jeigu būtų apimti visi testavimui skirti duomenys.



2.5.5 pav. Kelių segmentavimo, naudojant Masačusetso kelių duomenų rinkinį, rezultatų vaizdavimas [60]

2.5.2. Pastatų segmentavimas

Pastatų aptikimas ir atpažinimas aerofotografinėse nuotraukose yra labai svarbūs uždaviniai miestams planuoti, žemėlapiams sudaryti, pokyčiams nustatyti ir humanitarinei pagalbai organizuoti. Šiuos uždavinius apsunkina dideli pastatų išvaizdos skirtumai (pastatai pasižymi skirtingomis formomis, dydžiais, spalvomis ir kt.), kintančios apšvietimo sąlygos, šešėliai ir kt.[61].

2018 m. vykusio „SpaceNet“ organizuojamo pastatų aptikimo varžytinėse „U-Net“ parodė geriausius rezultatus [5]. Iššūkyje buvo naudojamas „SpaceNet“ duomenų rinkinys, kurį sudaro palydovinės nuotraukos, darytos Las Vegase, Paryžiuje, Šanchajuje ir Chartume. Daugelis pastatų buvo atpažinti vidutiniškai didesniu kaip 80 % tikslumu pagal IoU rodiklį. Varžybų nugalėtojas išskyrė tokius trūkumus kaip: mažesnis tikslumas mažesniems objektams, ypač, Šanchajuje ir Chartume; modelis labai tankiai išdėstytus namus traktuoja kaip vieną didelį pastatą [62].

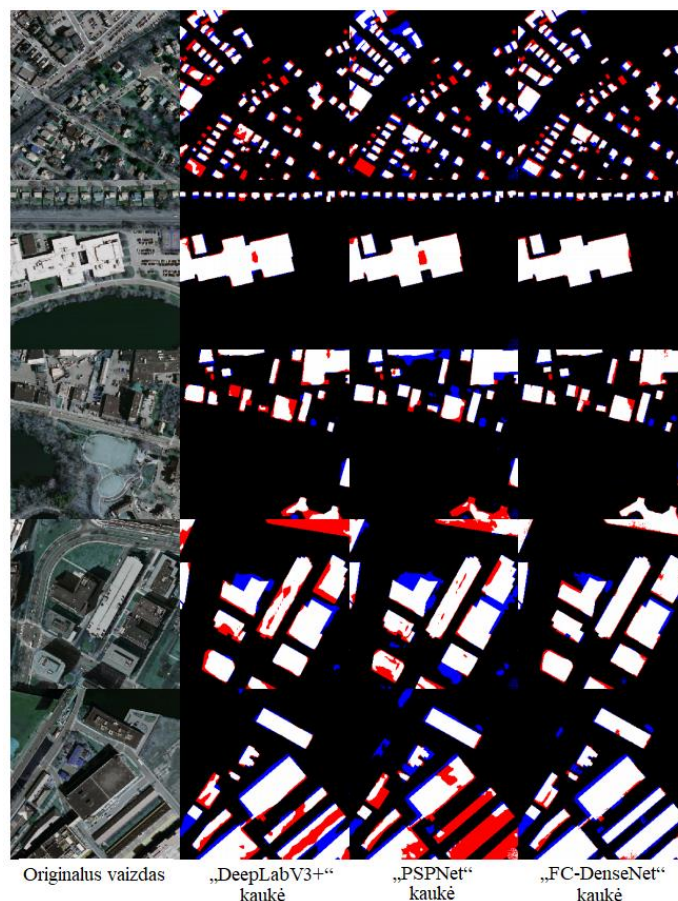
K. Sebastianas ir kt. (Sebastian et al., 2020a), atlikdami pastatų segmentavimo tyrimus, pristatė naują tikslo funkciją, sujungiančią priešpriešinę (angl. *Adversarial*) ir kryžminės entropijos funkcijas. Priešpriešinė tikslo funkcija paremta generatyvinių priešpriešinių tinklų metodika, kuri šiame darbe nėra aprašoma. Pasak autorių, tokia modifikacija padeda modeliui išmokti tiek vietinį, tiek globalųjį kontekstą. Eksperimentams atlikti buvo pasirinktas Masačusetso pastatų duomenų rinkinys. Duomenims papildyti buvo atliekamos horizontalaus ir vertikalios apvertimo, ir 90°, 180°, 270°

pasukimo operacijos. Autoriai tarpusavyje lygino „DeepLabV3+“, „PSPNet“ ir „FC-DenseNet“ segmentavimo modelių galimybes (žr. 2.5.6 lentelę). Pabrėžiama, kad „DeepLabV3+“ ir „PSPNet“ modeliams buvo priskirti svoriai, gauti iš „ImageNet“ duomenų rinkinio, o „FC-DenseNet“ modeliui priskirti atsitiktiniai pradiniai svoriai. Rezultatams įvertinti buvo taikomi modifikuoti mIoU ir „F1“ įverčiai, t. y. akcentuojamos tik dominančios srities pikseliai, atmetant didžiąją dalį fono.

2.5.6 lentelė. Pastatų segmentavimo rezultatai, naudojant Masačusetso pastatų duomenų rinkinį [61]

Metodas	mIoU, %	„F1“, %
„DeepLabV3+“	86,31	92,65
„PSPNet“	83,78	91,17
„FC-DenseNet“	91,55	95,59

2.5.6 paveiksle vaizduojamos tirtų modelių prognozuotos kaukės. Baltos, juodos, mėlynos ir raudonos spalvos pikseliai atitinkamai rodo TKTP (pastatai), TKNP (fonas), KKTP (klaidingai klasifikuoti kaip pastatai), KKNP (klaidingai klasifikuoti kaip fonas). Iš 2.5.6 lentelės matyti, kad „FC-DenseNet“ modelis, naudojant pasiūlyta tikslo funkciją, parodė geriausią rezultatą – 91,55 % modifikuotą mIoU vertę. Autoriai pabrėžė, kad atlikus didesnę duomenų papildymą, galima pasiekti dar geresnius rezultatus. Taikant naująją tikslo funkciją, trūkstamieji regionai užpildomi nuosekliau nei taikant tik kryžminės entropijos funkciją.



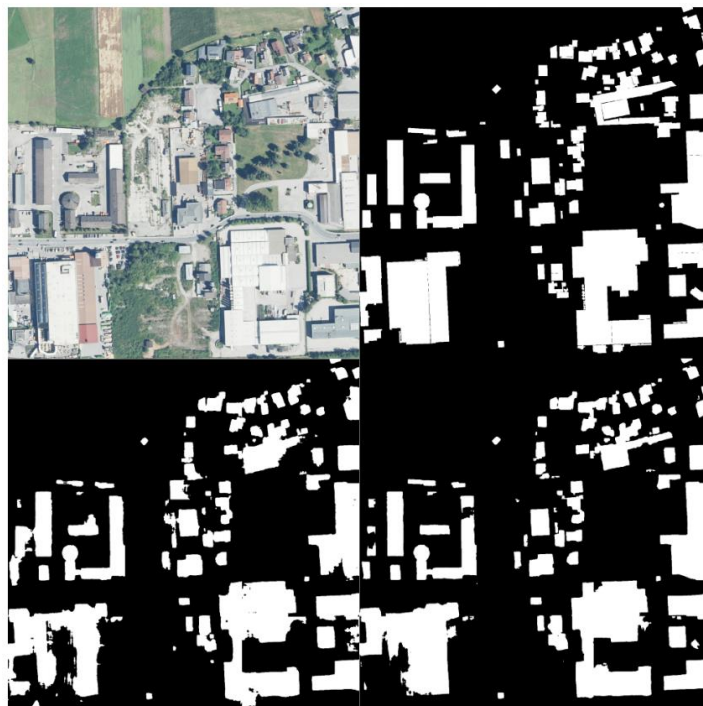
2.5.6 pav. Pastatų segmentavimo, naudojant Masačusetso pastatų duomenų rinkinį, rezultatų vaizdavimas [61]
AUTOMATINĖ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖ TAIKANT GILIOJO MOKYMOSI METODUS

K. Sebastianas ir kt. (Sebastian et al., 2020b) pristatė naują ir efektyvų būdą patobulinti pastatų segmentavimą aerofotografinėse nuotraukose. Enkoderio dalyje buvo įterptas konteksto erdvinio pasiskirstymo blokas – KEP. Naudojant KEP bloką yra išskiriama įvairaus mastelio kontekstinė informacija. Be to, fiksuojami ne tik erdviniai ryšiai, bet ir priklausomybės tarp skirtingų kanalų (angl. *Channel-wise Attention*). Šį kartą autoriai naudojo aukštos raiškos pastatų duomenų rinkinį „INRIA“. Tarpusavyje buvo lyginami modifikuotas VKT, „SegNet“ ir modifikuotas FPN. Pastarajam modeliui buvo pritaikomi skirtingi enkoderiai, apmokyti, naudojant „ImageNet“ duomenų rinkinį. Eksperimentų rezultatai pateikti 2.5.7 lentelėje.

2.5.7 lentelė. Pastatų segmentavimo rezultatai, naudojant „INRIA“ duomenų rinkinį [63]

Metodas	Enkoderis	IoU, %	PA, %
VKT + DP	-	64,67	94,42
„SegNet“	-	73	95,73
„FPN-KEP“	ResNet18	75,51	96,19
	ResNet101	77,03	96,47
	SE-ResNetXt-101	77,29	96,48

Matyti, kad geriausius rezultatus parodė modifikuotas FPN modelis, turintis „SE-ResNetXt-101“ enkoderį (77,29 % IoU). Kita vertus, autoriai pabrėžė, kad naudojant net tokį paprastą enkoderį kaip „ResNet18“, jų segmentavimo metodas užtikrina konkurencingus moderniems modeliams rezultatus. KEP bloko įtaka vizualiai parodyta 2.5.7 paveiksle.



2.5.7 pav. Pastatų segmentavimo, naudojant „INRIA“ duomenų rinkinį, rezultatų vaizdavimas. Viršuje yra originalus paveikslas ir jo tikroji kaukė, o apačioje modelio segmentavimo rezultatas be KEP bloko (kairėje) ir su KEP bloku (dešinėje) [63]

X. Li ir kt. (Li et al., 2021), atlikdami pastatų segmentavimo tyrimus, pristatė viena iš naujausių segmentavimo modelių – „PFNet“. Autoriai modifikavo FPN tinklą, naudodami savo sukurtą „PointFlow“ modulį, kuris kontekstinei informacijai gauti naudoja atraminius taškus [64]. Šiame darbe autorių minėtas modulis nėra detalizuojamas, o tik pateikiami galutiniai tyrimo rezultatai. Eksperimentams atlikti buvo pasirinkti Faihingenio prie Enco ir Potsdamo pastatų duomenų rinkiniai. Duomenims papildyti buvo naudotos horizontalaus ir vertikalaus apvertimo operacijos ir pasukimo atsitiktiniu kampu operacija. Tai pat buvo pasirinkta DKE tikslo funkcija. Rezultatams įvertinti buvo taikomi vidutiniai IoU ir „F1“ įverčiai. 2.5.8–2.5.9 lentelėse pateikiami tirtų modelių pastatų segmentavimo rezultatai.

2.5.8 lentelė. Pastatų segmentavimo rezultatai, naudojant Faihingenio prie Enco pastatų duomenų rinkinį [64]

Metodas	Enkoderis	mIoU, %	vidutinis „F1“ įvertis, %
VKT	„ResNet50“	64,2	75,9
„PSPNet“		65,1	76,8
„DeepLabV3+“		64,3	76
„PFNet“		70,4	81,9

2.5.9 lentelė. Pastatų segmentavimo rezultatai, naudojant Potsdamo pastatų duomenų rinkinį [64]

Metodas	Enkoderis	mIoU, %	vidutinis „F1“ įvertis, %
VKT	„ResNet50“	73,1	83,1
„PSPNet“		73,9	83,9
„DeepLabV3+“		74,1	83,9
„PFNet“		75,4	84,8

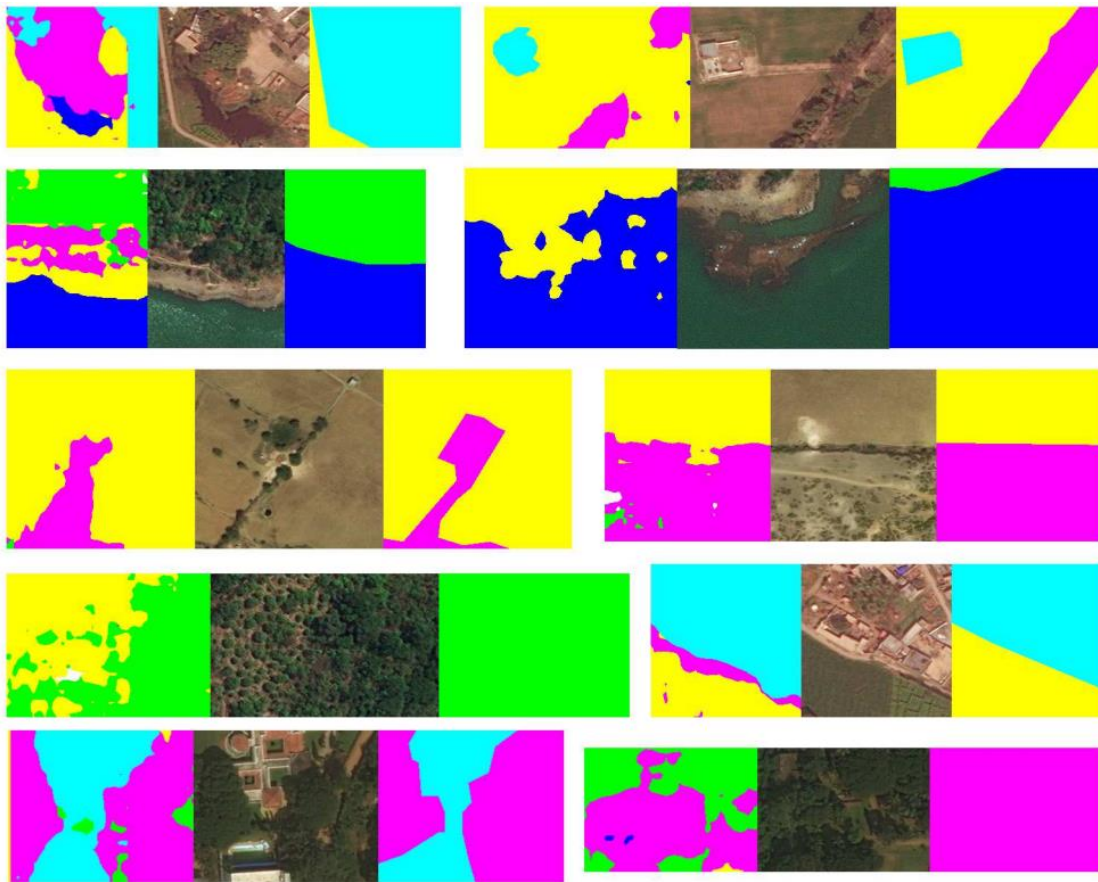
Lygindami „PFNet“ su kitų tyrėjų darbais, autoriai nustatė, kad „PFNet“ pasiekia geriausią spartos ir tikslumo kompromisą [64]. „PFNet“ pasiekė 70,4 % IoU ir 75,4 % IoU rodiklius, naudojant Faihingenio prie Enco ir Potsdamo duomenų rinkinius. Be to, tai yra vienas iš moderniausių segmentavimo modelių, kuris gali būti pritaikytas realaus laiko sistemoms.

2.6. Segmentavimo modelių tyrimas daugelio klasių segmentavimo atveju

Daugelio klasių segmentavimas yra vienas iš svarbiausių ir sudėtingiausių uždavinių kompiuterinės regos srityje. Automatinis žemės dangos suskirstymas į daugelį klasių padėtų efektyviau vystyti žemės ūkį, miškininkystę ir žemėtvarką [5].

I. Demir ir kt. (Demir et al. 2018) atliko tyrimus, sprenddami daugelio klasių segmentavimo uždavinį. Eksperimentams atlikti, buvo pritaikytas „DeepLab“ segmentavimo modelis, turintis „ResNet18“ enkoderį. Autoriai pasitelkė savo sukurtą duomenų rinkinį – „DeepGlobe“, kuriame

ekspertai sužymėjo septynias klases: miesto zoną, žemės ūkio zoną, miškus, augaliją, vandens telkinius, nederlinga žemę (kalnus, akmenis, dykumas, paplūdimius) ir kita. Duomenims papildyti buvo naudota tik pasukimo transformacija. Siekiant išvengti didelio klasių disbalanso, kiekvienai klasei buvo suteiktas tam tikras svorio koeficientas. 2.6.1 paveiksle pavaizduoti segmentavimo rezultatai. Žydra spalva nuspalvinta miesto zona; geltona spalva nuspalvinta žemės ūkio zona; violetine spalva nuspalvinta augalija; žalia spalva nuspalvinti miškai; mėlyna spalva nuspalvinti vandens telkiniai. Geriausi rezultatai parodė 43,3 % mIoU rodiklį [5]. Lyginant daugelio klasių segmentavimo rezultatus su dviejų klasių segmentavimo rezultatais, matyti, kad šioje srityje dar trūksta tikslumo.



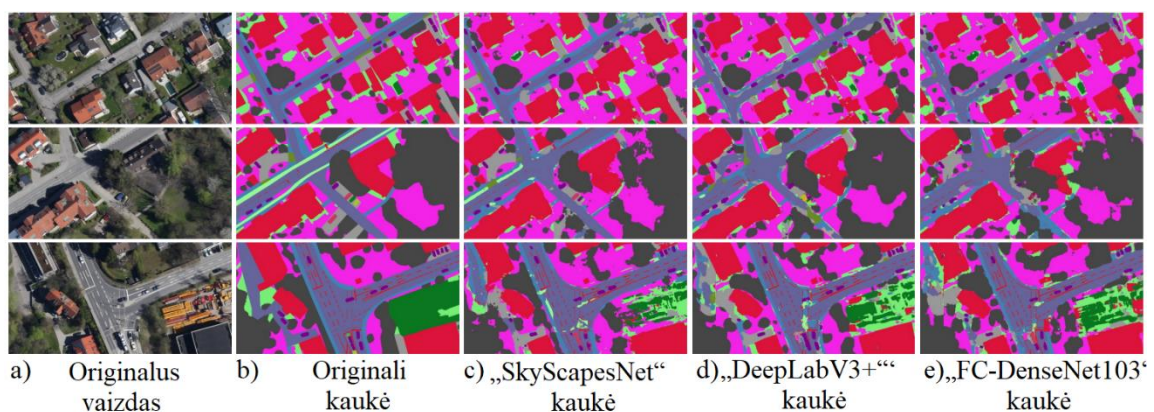
2.6.1 pav. 7 klasių segmentavimo, naudojant „DeepLab“ modelį ir „DeepGlobe“ duomenų rinkinį, rezultatų vaizdavimas [5]

S. M. Azimi ir kt. (Azimi et al., 2020) atliko išsamius daugelio klasių segmentavimo tyrimus. Autoriai naudojo „SkyScapes-Dense“ duomenų rinkinį, kuriame buvo išskirta 20 pagrindinių klasių (pvz., keliai, augmenija, pastatai, stovėjimo aikštelės ir t. t.). Duomenims papildyti buvo taikomos horizontalaus ir vertikalios apvertimo operacijos. Autoriai pristatė naują segmentavimo modelį, kurį pavadino „SkyScapesNet“ ir palygino rezultatus su „FCN-8s“, „U-Net“, „PSPNet“, „RefineNet“ ir „DeepLabV3+“ rezultatais. Buvo taikyta KE ir „Dice“ tikslo funkcijų modifikacija. Rezultatams įvertinti buvo pasitelktos mIoU, preciziškumo ir specifiškumo metrikos. Tyrimų rezultatai pateikiami 2.6.1 lentelėje. Lyginamųjų modelių prognozuotos kaukės parodytos 2.6.2 paveiksle.

2.6.1 lentelė. 20 klasių segmentavimo rezultatai, naudojant „SkyScapes-Dense“ duomenų rinkinį [65]

Metodas	Enkoderis	mIoU, %	Precision, %	Recall, %
„FCN-8s“	„ResNet50“	33,06	65,01	40,78
„U-Net“	-	14,15	22,87	21,88
„PSPNet“	„ResNet101“	30,44	43,63	40,48
„RefineNet“	„ResNet152“	36,39	52,17	46,12
„DeepLabv3+“	„Xception65“	38,20	55,34	47,97
„SkyScapesNet“	-	40,13	65,93	47,85

Matyti, kad vidutinis geriausio modelio tikslumas nesiekia 50 % IoU įverčio. Tačiau net ir su tokiu mažu tikslumu, segmentavimo kaukės atrodo labai neblogai (žr. 2.6.2 pav.).

**2.6.2 pav.** 20 klasių segmentavimo, naudojant „SkyScapes“ duomenų rinkinį, rezultatų vaizdavimas [65]

M. Hefelsas ir J. Vanšorenas (Heffels & Vanschoren, 2020), akcentuodami nepakankamą daugelio klasių segmentavimo tyrimų skaičių nuotolinio aptikimo srityje, sukūrė viešai prieinamą duomenų rinkinį „DroneDeploy“ ir išbandė „DeepLabV3+“, „U-Net“ modelių galimybes. Duomenų rinkinyje buvo išskirtos šešios klasės: pastatai, augalija, vandens telkiniai, žemė, automobiliai ir pašaliniai objektai. Eksperimentams atlikti buvo taikytos KE ir ŽTF tikslo funkcijos. Rezultatams įvertinti naudotas mIoU įvertis. 2.6.2 lentelėje pateikiami 6 klasių segmentavimo rezultatai.

2.6.2 lentelė. 6 klasių segmentavimo rezultatai, naudojant „DroneDeploy“ duomenų rinkinį [66]

Metodas	Enkoderis	Naudota biblioteka	Tikslo funkcija	mIoU, %
„DeepLabV3+“	„Xception65“	-	KE	52,5
	„ResNetV1-50“		KE	48
„U-Net“	„ResNet50“	„Keras“	ŽTF	39,8
			KE	42,5
		„fastai“	ŽTF	30,2
			KE	29,4



2.6.3 pav. Automobilių ir augalijos segmentavimo pavyzdys, taikant „DeepLabV3+“ modelį [66]

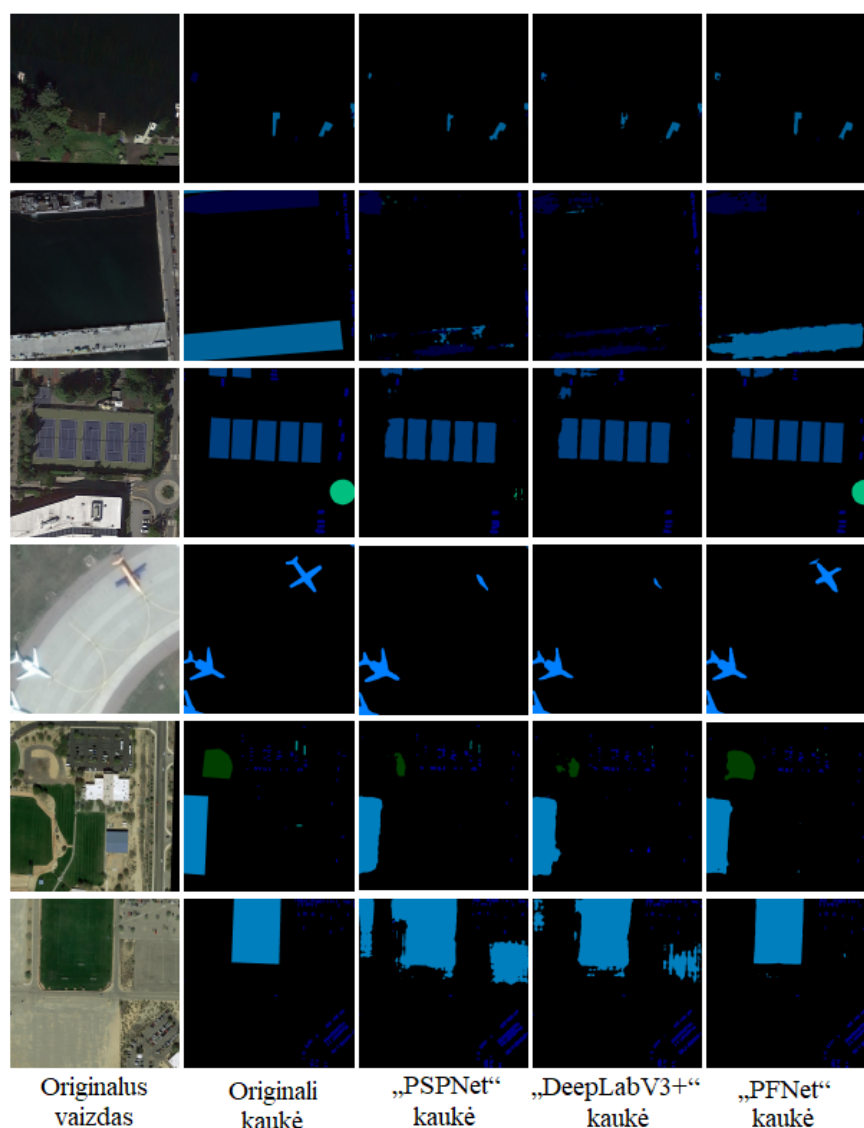
Iš lentelės matyti, kad naudojant skirtingas bibliotekas, segmentavimo rezultatai skiriasi. Autoriai pabrėžė, kad židinio tikslo funkcija nepagerino modelių tikslumo. Geriausius rezultatus rodantis modelis, sugeba išskirti mažus objektus tokius kaip automobiliai, tačiau jie atitinkamai segmentuojami prasčiausiu tikslumu.

X. Li su bendraautoriais (Li et al., 2021) atliko daugelio klasių segmentavimo tyrimus, naudodami „iSAID“ duomenų rinkinį, ir palygino savo sukurtą „PFNet“ tinklo rezultatus su „DenseASPP“, „DeepLabV3+“ ir „PSPNet“ modelių segmentavimo rezultatais. „iSAID“ duomenų rinkinį sudaro 2,806 aukštos raiškos nuotraukos, kuriose buvo sužymėta 14 klasių. Eksperimentams atlikti buvo taikyta KE tikslo funkcija, o rezultatams įvertinti naudotas mIoU įvertis. Tyrimų rezultatai pateikti 2.6.3 lentelėje.

2.6.3 lentelė. 14 klasių segmentavimo rezultatai, naudojant „iSAID“ duomenų rinkinį [64]

Metodas	Enkoderis	mIoU, %
„DenseASPP“	„ResNet50“	57,3
„DeepLabV3+“		61,2
„PSPNet“		60,3
„PFNet“		66,9

Matyti, kad autorių sukurtas segmentavimo modelis žymiai pagerino ankstesnius daugelio klasių segmentavimo rezultatus, ir pasiekė ~67 % mIoU įvertį. Pabrėžiama, kad „PFNet“ sugeba išskirti įvairaus dydžio objektus išlaikydamas jų formą. 2.6.4 paveiksle pateikta „PFNet“ segmentavimo kaukė labiausiai atitinką originalią kaukę.



2.6.4 pav. 14 klasių segmentavimo, naudojant „iSAID“ duomenų rinkinį, rezultatų vaizdavimas [64]

2.7. Trumpas skyriaus apibendrinimas

Šiame skyriuje buvo aptarti aerofotografinių nuotraukų segmentavimo iššūkiai ir modernūs giliojo mokymosi metodai, tinkantys vaizdams segmentuoti. Išnagrinėtos įvairios tikslo funkcijos bei segmentavimo rezultatų įvertinimo metrikos. Remiantis kitų tyrėjų eksperimentais, ištirtos kelių, pastatų ir daugelio klasių segmentavimo galimybės. Darytina išvada, kad giliojo mokymosi metodai yra labai perspektyvūs aerofotografiniams vaizdams analizuoti. Pirma, segmentavimo modeliai geba pakankamai aukštu tikslumu automatiškai išskirti kelius iš vaizdo fono. Galima išskirti „PP-LinkNet“ modelį, kuris pasiekė 70,53 % IoU vertę, naudojant „DeepGlobe“ duomenų rinkinį (4 % pranašesnis už „DeepGlobe“ varžybų nugalėtoją). Šis modelis geriau nei kiti išskiria siauresnius ir trumpesnius kelius. Be to, pastebima mažiau trūkių. „PP-LinkNet“ modelis parodė geras apibendrinimo galimybes. Antra, segmentavimo modeliai, geba išskirti įvairaus dydžio pastatus, tačiau sunkiai susitvarko išlaikydami tikrąją pastatų formą. Pastatų segmentavimo atveju galima išskirti vieną iš moderniausių metodų „PFNet“. Šis modelis, lyginant su kitais, pasiekia geriausią tikslumo ir

AUTOMATINĖ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖ TAIKANT GILIOJO MOKYMOSI METODUS

skaičiavimo sąnaudų santykį. Naudojant Faihingo prie Enco ir Potsdamo duomenų rinkinius, buvo pasiekti 70,4 % IoU ir 75,4 % IoU įverčiai. „PFNet“ prognozuotos kaukės atrodo natūraliausiai, nes yra išlaikomi aštrūs pastatų kampai, Kiti modeliai yra linkę užapvalinti kampus. Pagaliau, giliojo mokymosi metodai gali padėti spręsti daugelio klasių segmentavimo problemas. Šiuo atveju galima taip pat išskirti „PFNet“ modelį, kuris, naudojant „iSAID“ duomenų rinkinį, rodo daugiau kaip 5 % geresnius rezultatus pagal mIoU įvertį, lyginant su kitais modeliais. Nepaisant to, šiuolaikiniams segmentavimo modeliams vis dar trūksta tikslumo, tačiau jie žymiai palengvina varginantį duomenų žymėjimo procesą Segmentavimo rezultatai priklauso nuo duomenų kiekio, taikomų tikslo funkcijų, žmogaus sužymėtų kaukių kokybės, duomenų papildymo ir kt.

3. DUOMENŲ PARUOŠIMAS AUTOMATINĖS AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZĖS METODŲ EKSPERIMENTINIAM TYRIMUI

Šiame skyriuje pateikiami viešai prieinami duomenų rinkiniai, kurie buvo panaudoti atrinktiems aerofotografinių nuotraukų analizės metodams ištirti. Nagrinėjami duomenų rinkiniai taikomi dvejetainio ir daugelio klasių segmentavimo uždaviniams spręsti. Duomenų atrankos metu buvo atsižvelgiama į galimybę palyginti eksperimentų rezultatus su kitų tyrėjų gautais rezultatais. Kompiuterinės regos srityje duomenų rinkiniai įprastai yra suskirstomi į mokymo, validacijos ir testavimo rinkinius, siekiant išvengti vadinamojo persimokymo (angl. *Overfitting*) ir pasirinkimo šališkumo (angl. *Selection Bias*) problemų [67]. Mokymo rinkinys yra viso duomenų rinkinio poaibis, skirtas modeliui mokytis. Modelis mokydamasis iš šių duomenų, optimizuoja vadinamuosius hiperparametrus (kintamuosius, kurie nusako modelio konfigūraciją). Validacijos duomenų rinkinys reikalingas nešališkai įvertinti modelį, derinant modelio hiperparametrus. Testavimo rinkinys – modeliui nematyti duomenys, kurie naudojami, siekiant nešališkai įvertinti validacijos metu atrinktą galutinį modelį. Vykstant įvairioms varžytuvėms, testavimui skirti duomenys tampa prieinami tik tada, kai konkursas artėja prie pabaigos, o nugalėtoją nulemia modelio rezultatas, naudojant testavimui skirtus duomenis [67]. Visas duomenų rinkinys į poaibius suskirstomas priklausomai nuo duomenų kiekio ir jų tipo. Jei duomenų rinkinio dydis yra nuo 100 iki 10 000, jį įprasta dalinti santykiu 60:20:20. Tai yra 60 % duomenų skirta mokymo poaibiui ir po 20 % – validacijos ir testavimo poaibiams. Jeigu duomenų yra daugiau nei 1 mln., rinkinys dalinamas santykiu 98:1:1 arba 99:0.5:0.5 [67]. Kita vertus, čia griežtų dalybos taisyklių nėra, svarbu, kad, modelis nebūtų šališkas tendencijoms ir nepradėtų įsiminti tam tikrų požymių. Pažymėtina, kad mokant ir validuojant modelį duomenų statistinis pasiskirstymas turi būti kuo panašesnis. Turint mažai duomenų, jų kiekį dirbtinai galima padidinti, taikant duomenų papildymo metodus, kurie plačiau aprašomi 3.7 skyriuje. Lyginant kelių ir pastatų semantinę informaciją, galima sakyti, kad pastatų struktūra yra sudėtingesnė, nes jie gali būti įvairaus dydžio, formų ir aštrių kampų.

3.1. Masačusetso kelių duomenų rinkinys

Masačusetso kelių duomenų rinkinį sudaro 1 171 spalvota aerofotografinė nuotrauka, apimanti 2,25 kvadratinį kilometrą Masačusetso valstijos žemės plotą. Kiekvienos nuotraukos raiška yra 1500×1500 pikselių, kur vienas vaizdo taškas atitinka 1 m² žemės ploto [68]. Masačusetso kelių duomenų rinkinys iš viso apima daugiau kaip 2 600 km² įvairių miesto, priemiesčio ir kaimo vietovių, įskaitant augaliją, pastatus, kelius, upes ir transporto priemones, tad dėl skirtingos semantinės informacijos šis rinkinys taikomas moderniems segmentavimo modeliams tirti. Kaukės (angl. *Masks*) nuotraukoms buvo padarytos naudojant atvirąjį žemėlapių „OpenStreetMap“, vektorines kelių

centrines linijas pavertus pikseliais [68]. 3.1.1 paveiksle pateikiami aerofotografinės nuotraukos ir atitinkamos kaukės pavyzdžiai. Pradinį duomenų rinkinį sudaro 3 dalys: 1 108 nuotraukų, skirtų modeliui apmokyti, 14 nuotraukų, skirtų modeliui validuoti, ir 49 nuotraukos, skirtos modeliui testuoti. Pažymėtina, kad dalis apmokymui skirtų nuotraukų turi baltus lopus, kurie turi būti pašalinti. Atmetus netinkamas aerofotografinės nuotraukas visas rinkinys sumažėjo iki 749 nuotraukų. Dėl skaičiavimo išteklių apribojimų aerofotografinės nuotraukos papildomai padalijamos į mažesnes dalis. Šiame darbe eksperimentams atlikti buvo sudaryti nauji rinkiniai. Viena originalaus dydžio vaizdai buvo padalinti į 384×384 pikselių raiškos vaizdus, o kitame į 512×512 pikselių raiškos vaizdus.



3.1.1 pav. Masačusetso kelių duomenų rinkinio aerofotografinė nuotrauka ir jos kaukė [68]

3.2. „DeepGlobe“ kelių duomenų rinkinys

Šį duomenų rinkinį sudaro 8 570 spalvotų aerofotografinių nuotraukų, apimančių 2 220 kvadratinį kilometrų Tailando, Indonezijos ir Indijos teritorijų plotą. Kiekvienos nuotraukos raiška yra 1024×1024 vaizdo taškų, kur vienas pikselis atitinka 50 cm^2 žemės plotą (dvigubai didesnė erdvinė skyra, lyginant su Masačusetso kelių duomenų rinkiniu) [5]. 3.2.1 paveiksle pateikiama aerofotografinė nuotrauka ir ją atitinkanti kaukė. Pažymima, kad duomenų rinkinį sudaro vienoda dalis skirtingo tipo kelių (asfaltuotų, neasfaltuotų ir t. t.), esančių kaimo ir miesto vietovėse. Kaukės nuotraukoms buvo padarytos ekspertų, naudojant atvirojo kodo informacinę sistemą „QGIS“. Kitaip nei Masačusetso duomenų rinkinyje, čia kaukės buvo daromos pikselių lygiu, o ne pagal centrinę liniją, todėl yra tikslesnės [5]. Tiesa, rinkinyje atitinkamas dvejetainis kaukės turi tik 6 226 nuotraukos. Miestų sandara, apšvietimo sąlygos, kelių tankumas ir gatvių tinklų struktūra yra labai įvairi, todėl šis duomenų rinkinys naudojamas ne tik skirtingiems modeliams įvertinti, bet ir lyginamajai analizei atlikti. Šio darbo eksperimentams atlikti buvo panaudotos tik tos nuotraukos, kurios turi kaukių. Padalijus nuotraukas į mažesnes dalis, kurias sudaro 512×512 pikselių raiškos vaizdai, buvo gauta nauja duomenų aibė. Mokymo poaibį sudaro 20 904 nuotraukos, o validacijos ir testavimo poaibius – po 2 000 nuotraukų.



3.2.1 pav. „DeepGlobe“ kelių duomenų rinkinio aerofotografinė nuotrauka ir kaukė [5]

3.3. Masačusetso pastatų duomenų rinkinys

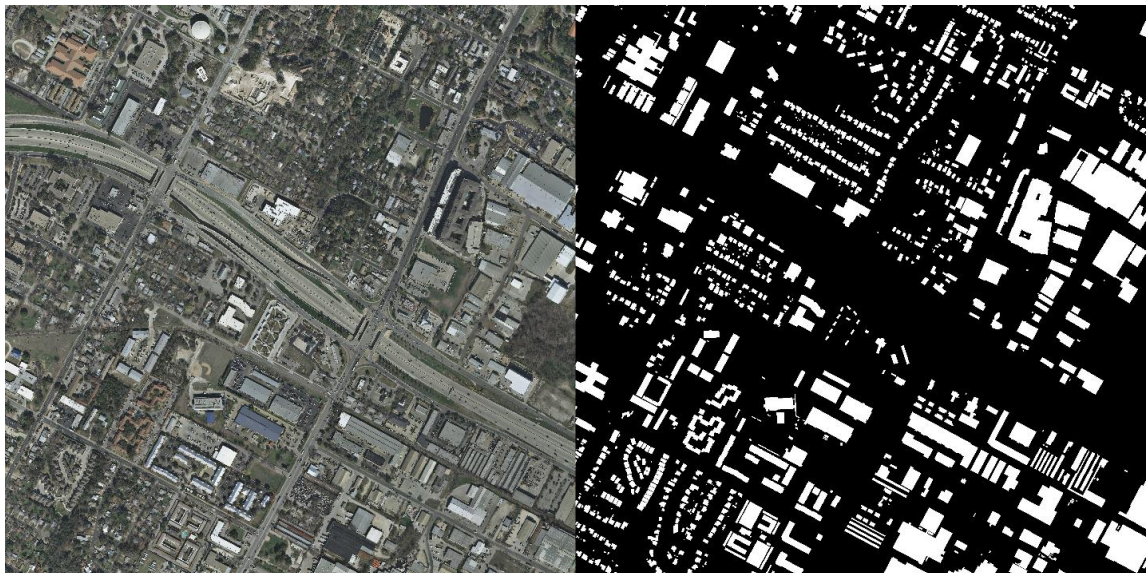
Masačusetso pastatų duomenų rinkinį sudaro 151 Bostono miesto spalvota aerofotografinė nuotrauka, apimanti 2,25 kvadratinį kilometrų plotą. Kiekvienos nuotraukos raiška yra 1500×1500 pikselių, o vienas pikselis atitinka 1 m^2 žemės ploto [68]. Duomenų rinkinys užima apie 340 km^2 plotą. Kaukės, kaip ir Masačusetso kelių duomenų rinkinyje, buvo padarytos naudojant atvirąjį žemėlapi „OpenStreetMap“. Duomenų rinkinys apima Bostono miesto ir priemiesčių teritorijas, kuriose yra įvairaus dydžio pastatai, įskaitant nuosavus namus ir garažus. Duomenų rinkinio pavyzdys pateiktas 3.3.1 paveiksle. Kadangi, pradinių duomenų yra nemažai, jie buvo padalinti į mažesnes dalis, kurių vienos nuotraukos raiška sumažinta iki 384×384 pikselių. Taigi, eksperimentinis duomenų rinkinys susideda iš 2 192 mokymui skirtų duomenų, 64 validacijai skirtų duomenų ir 160 testavimui skirtų duomenų.



3.3.1 pav. Masačusetso pastatų duomenų rinkinio aerofotografinė nuotrauka ir kaukė [68]

3.4. „INRIA“ pastatų duomenų rinkinys

„INRIA Aerial Image Labeling Dataset“ – tai aukštos raiškos duomenų rinkinys, kurį sudaro 360 spalvotų aerofotografinių nuotraukų, apimančių 810 kvadratinį kilometrų žemės plotą. Kiekvienos nuotraukos raiška yra 5000×5000 pikselių, o vienas pikselis atitinka 30 cm^2 žemės plotą (daugiau kaip 3 kartus didesnė erdvinė skyra, lyginant su Masačusetso pastatų duomenų rinkiniu). Apskritai, „INRIA“ rinkinyje yra užfiksuoti dešimt skirtingų Europos ir JAV miestų, iš kurių 5 miestai (Ostinas, Čikaga, Kitsapo apskritis, Vakarų Tirolis, Viena) yra skirti tinklui mokyti [63]. Tikėtina, kad įvairesni duomenys suteikia modeliui geresnių apibendrinimo (angl. *Generalization*) galimybių, tai reiškia, kad apmokytas modelis galės tiksliau segmentuoti dar nematytą aerofotografinę nuotrauką. Pažymima, kad kaukes turi tik mokymo poaibis, sudarytas iš 180 aerofotografinių nuotraukų, todėl ir šį duomenų rinkinį reikėjo pertvarkyti. Pagaliau buvo sudarytas naujas duomenų rinkinys, susidedantis iš 18 000 (512×512 pikselių raiškos) nuotraukų, padalintų į poaibius pagal proporciją 80:10:10.



3.4.1 pav. „INRIA“ pastatų duomenų rinkinio aerofotografinė nuotrauka ir kaukė [69]

3.5. Lietuvos keliai

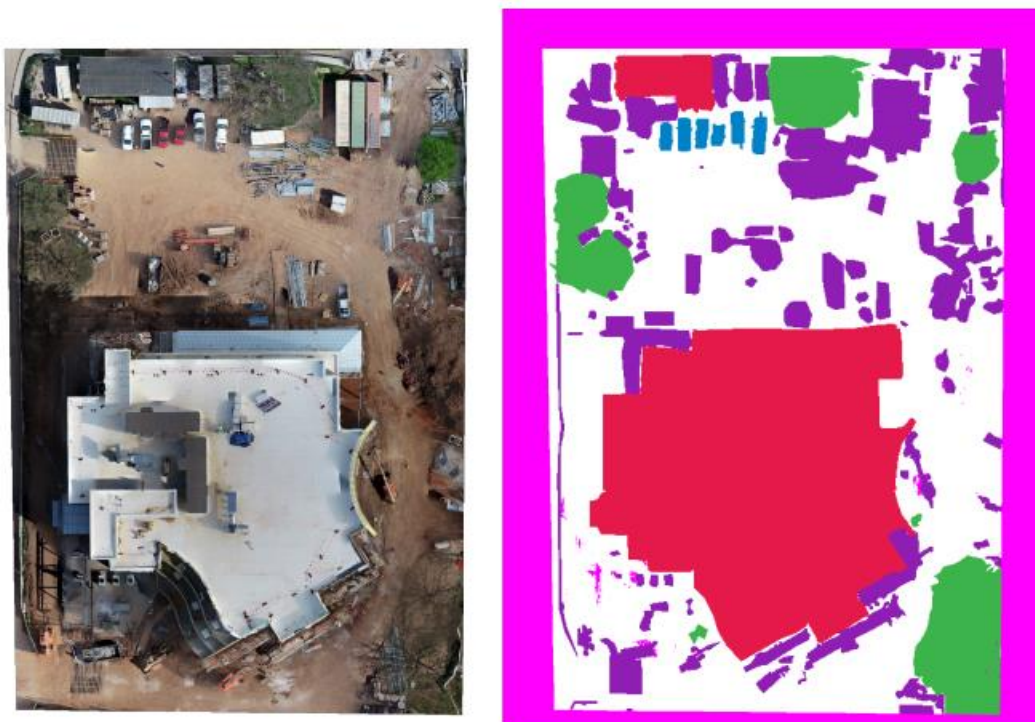
Lietuvos kelių duomenų rinkinys buvo specialiai sukurtas papildomiems tyrimams atlikti. Jį sudaro 100 spalvotų mažos raiškos aerofotografinių nuotraukų, apimančių Vilnių, Kauną, Klaipėdą, Panevėžį, Trakus, Elektrėnus, Rumšiškės, Žiežmarių, Kaišiadoris, Prienuos ir Birštoną. Kiekvienos nuotraukos raiška yra 1000×1000 pikselių. Vaizdai buvo gauti darant ekranvaizdžius iš interaktyvios lietuviškos internetinių žemėlapių svetainės *Maps.lt*, parinkus 100 m. skalę, o kaukės iš svetainės *Snazzymaps.com*, parinkus tik kelių rodymo stilių ir sutapatinus vaizdo mastelius. Duomenys papildomai buvo padalinti į mažesnes dalis, kur kiekviena nuotrauka yra 320×320 raiškos. Mokymo poaibį sudaro 1 280 aerofotografinių nuotraukų, o validacijos ir testavimo poaibius – 160 nuotraukų.

3.6. „DroneDeploy“ duomenų rinkinys

Duomenų rinkinį sudaro 55 spalvotos aerofotografinės nuotraukos. Kitaip nei aptarti duomenų rinkiniai, pastarasis buvo darytas naudojant bepilotės skraidykles [70]. Aeronuotraukos yra skirtingos raiškos, bet vienodos 0,1 km² erdvinės skyros. Vienas pikselis atitinka 10 cm² žemės plotą. Nuotraukose buvo išskirtos šešios klasės: pastatai, kiti objektai, augalija, vandens telkiniai, žemė ir automobiliai. Be to, buvo papildomai išskirta klasė „Nepaisyti“, sujungianti nesužymėtus pikselius ir nuotraukų ribas. Pradiniai vaizdai yra per dideli apdoroti vienu metu, todėl kiekviena nuotrauka buvo padalinta į mažesnes 512×512 raiškos nepersidengiančias dalis, atmetant „Nepaisyti“ klasei priklausančius pikselius. Padalinus nuotraukas, mokymo poaibį sudarė 1717, validavimo poaibį 281, o testavimo poaibį 987 nuotraukos. 3.6.1 lentelėje parodoma, kiek pikselių priklauso kiekvienai klasei. 3.6.1 paveiksle parodytas daugelio klasių kaukės pavyzdys.

3.6.1 lentelė. Pikselių pasiskirstymas sužymėtų klasių „DroneDeploy“ duomenų rinkinyje [70]

Eil. Nr.	Klasė	Spalvos kodas	Pikselių skaičius, %
1	Pastatai	Raudona	5,6
2	Kiti objektai	Violetinė	2
3	Augalija	Žalia	10,43
4	Vandens telkiniai	Oranžinė	1,2
5	Žemė	Balta	37,7
6	Automobiliai	Mėlyna	0,38
7	„Nepaisyti“	Purpurinė	42,7



3.6.1 pav. „DroneDeploy“ duomenų rinkinio aerofotografinė nuotrauka ir kaukė [70]

3.7. Duomenų papildymas

Yra žinoma, kad didesnis duomenų kiekis gerina segmentavimo modelio rezultatą ir jo apibendrinimo galimybes [71]. Kita vertus, turint nedidelį duomenų rinkinį, giliojo mokymosi metodai gali rodyti blogesnius rezultatus už kai kurių tradicinių segmentavimo metodų rezultatus. Praktikoje duomenis sunku surinkti, o duomenų žymėjimas taip pat yra daug laiko reikalaujantis ir brangus darbas, todėl labai svarbų vaidmenį atlieka duomenų papildymas (angl. *Data Augmentation*) [71]. Kompiuterinės regos srityje duomenų papildymas – dirbtinis duomenų kiekio didinimas, taikant įvairias transformacijas. Tai taip pat yra vienas iš reguliarizavimo metodų, padedančiu kovoti su tokiomis problemomis kaip persimokymas ir t. t. Tiesa, atliekant duomenų papildymą, tie patys veiksmai atliekami su atitinkamų vaizdų kaukėmis. 3.7.1 lentelėje pateikiamos pagrindinės duomenų papildymo technikos ir trumpi jų aprašymai.

3.7.1 lentelė. Pagrindinės duomenų papildymo technikos [72]

Transformacijos tipas	Aprašymas
Mastelio keitimas	Vaizdas yra didinamas, mažinamas, praplečiamas arba suspaudžiamas
Apkarpymas	Išrenkama pasirinkto dydžio vaizdo dalis
Apvertimas	Vaizdas apverčiamas horizontaliai arba vertikalčiai
Užpildymas	Vaizdas praplečiamas, sudarant dirbtinį rėmelį
Pasukimas	Vaizdas pasukamas atsitiktiniu arba fiksuotu kampu
Postūmis	Objektai vaizde pastumiami koordinačių sistemoje
Afininė transformacija	Vaizdas geometriškai iškreipiamas
Ryškumo keitimas	Vaizdas tamsinamas arba šviesinamas
Kontrasto keitimas	Keičiamas santykis tarp tamsiausių ir ryškiausių vaizdo sričių
Sodrumo keitimas	Keičiamas spalvų intensyvumas
Atspalvio suteikimas	Keičiamos spalvos
Aštrinimas	Pabrėžiamos vaizdo detalės
Glotninimas	Vaizdas yra suliejamas

Aptarti duomenų papildymo metodai dažnai taikomi kartu, pvz., vaizdas pasukamas, paryškinamas padidinamas ir apkarpomas. Pabrėžiama, kad duomenų papildymas atliekamas tik mokymui skirtų duomenų rinkinyje [72].

3.8. Trumpas skyriaus apibendrinimas

Šiame skyriuje buvo atrinkti viešai prieinami, eksperimentams tinkantys, aerofotografinių nuotraukų duomenų rinkiniai. Aptarti duomenų papildymo metodai. Buvo išsiaiškinta, kad kompiuterinės regos srityje duomenų rinkiniai įprastai yra suskirstomi į mokymo, validacijos ir testavimo rinkinius, siekiant išvengti vadinamojo persimokymo ir pasirinkimo šališkumo problemų. Be to, išsiaiškinta, kad paprastai pradiniai vaizdai yra per dideli apdoroti vienu metu, todėl kiekvieną nuotrauką, atsižvelgiant į kompiuterinius išteklius, reikia padalinti į mažesnės raiškos dalis. Matyti, kad duomenų rinkiniai gali būti įvairios raiškos ir erdvinės skyros, tai turi įtakos segmentavimo rezultatams. Nustatyta, kad duomenų kiekis gali būti padidintas, taikant duomenų papildymo metodus. Dažnai taikoma įvairių duomenų papildymo technikų kombinacija. Duomenys daro didžiausią įtaką giliojo mokymosi metodų segmentavimo rezultatams.

4. GILIOJO MOKYMOŠI METODŲ, SKIRTŲ AEROFOTOGRAFINIŲ NUOTRAUKŲ ANALIZEI, TYRIMO REZULTATAI

Šiame skyriuje aprašomi tyrimams naudoti kompiuteriniai ištekliai ir programinė įranga. Pateikiami atliktų dvejetainio ir daugelio klasių segmentavimo eksperimentų, naudojant 2-ajame skyriuje ištirtus segmentavimo modelius ir 3-ajame skyriuje aprašytus duomenų rinkinius, rezultatai.

4.1. Programinė įranga ir kompiuteriniai ištekliai

Šiame darbe visi praktiniai tyrimai buvo atlikti, naudojant nemokamas debesijos paslaugas (angl. *Cloud Computing Services*) „Google Colab“ ir „Kaggle“. Šios platformos yra paremtos atvirojo kodo projektu „Jupyter“, kuriame naudojama „Python“ programavimo kalba. „Colab“ leidžia vartotojams naršyklėje rašyti ir vykdyti programos kodą. Tiesa, kompiuteriniai ištekliai nėra garantuoti ir neriboti, o naudojimo laikas kartais svyruoja. Tai būtina, kad „Colab“ galėtų nemokamai teikti išteklius visiems vartotojams ir išlaikytų lankstumą. Vartotojui priskiriamas vienas grafikos procesorius iš dviejų galimų („Nvidia Tesla K80“, „Nvidia Tesla T4“). Dėl šių priežasčių gali kisti tyrimų rezultatai. „Kaggle“ leidžia vartotojams naudotis viešais duomenų rinkiniais, tirti ir kurti giliojo mokymosi modelius internetinėje aplinkoje, dirbti su kitais tyrėjais ir mašininio mokymosi inžinieriais, dalyvauti konkursuose ir t. t. „Kaggle“ suteikia nemokamą prieigą prie „Nvidia Tesla P100“ grafikos procesoriaus. Vartotojui leidžiama naudoti suteiktus išteklius apie 30 valandų per savaitę. 4.1.1 lentelėje pateikiamas GP techninių charakteristikų palyginimas. Didesni skaičiai lemia aukštesnę spartą. Šiuolaikiniai gilieji dirbtinių neuronų tinklai turi labai daug parametrų, kuriuos reikia optimizuoti. Tam reikia daug skaičiavimo išteklių.

4.1.1 lentelė. GP techninių charakteristikų palyginimas

Parametras	„Tesla K80“	„Tesla T4“	„Tesla P100“
„CUDA“ branduoliai	2496	2560	3584
Taktinis dažnis, MHz	562	585	1190
Atminties dydis, GB	12	16	16
Atminties sparta, GB/s	240,6	320	732
Slankiojo kablelio operacijų per sekundę skaičius	$\sim 1,87 \cdot 10^{12}$	$\sim 0,25 \cdot 10^{12}$	$4,7 \cdot 10^{12}$

Eksperimentams atlikti buvo naudojamos aukštojo lygio sistemos „Keras“ ir „PyTorch“. Segmentavimo modeliams kurti buvo pasitelktos „segmentation_models“ (skirta „Keras“ sistemai) ir „segmentation_models_pytorch“ (skirta „PyTorch“ sistemai) bibliotekos [73, 74]. Duomenims papildyti buvo naudojama „albumentations“ biblioteka [75]. Pabrėžiama, kad visų eksperimentų metu buvo atliekamos tokios duomenų papildymo operacijos:

- horizontalus apvertimas (50 % tikimybė),
- vertikalus apvertimas (50 % tikimybė),
- postūmis + mastelio keitimas + pasukimas (maksimalus postūmis 10 %, mastelis keičiamas iki 50 %, pasukimas atliekamas atsitiktiniu kampu),
- pridedamas Gauso triukšmas (iki 20 %),
- iškreipimas (iki 50 %),
- ryškumo ir kontrasto keitimas (atsitiktinė tvarka),
- aštrinimas ir glotninimas (atsitiktinė tvarka),
- spalvų ir atspalvių keitimas (atsitiktinė tvarka).

Tikslo funkcijoms optimizuoti buvo pasirinktas „Adam“ optimizatorius.

4.2. Dvejetainio segmentavimo rezultatai

Keliams segmentuoti buvo naudojami pertvarkyti Masačusetso, „DeepGlobe“ ir Lietuvos kelių duomenų rinkiniai, o pastatams – Masačusetso ir „INRIA“ pastatų duomenų rinkiniai. Šis tyrimas naudingas tuo, kad galima palyginti įvairiomis sąlygomis darytų nuotraukų segmentavimo galimybes. Visgi, didžiausias efektyvaus segmentavimo modelio privalumas – geros apibendrinimo galimybės. Dvejetainio segmentavimo eksperimentams atlikti nebuvo naudojamos papildomos apdorojimo priemonės, turinčios įtaką galutiniams rezultatams.

4.2.1. Kelių segmentavimas

Pirmojo eksperimento metu, naudojant „Keras“ sistemą ir pertvarkytą Masačusetso kelių duomenų rinkinį, buvo tiriami „U-Net“, FPN, „PSPNet“ ir „LinkNet“ segmentavimo modeliai. Remiantis [46, 76], buvo pasirinktas „EfficientNetB3“ enkoderis. Modelių įvestyje paduodamos 384×384 pikselių dydžio nuotraukos. Pagal [56], buvo pasirinkta kombinuotoji tikslo funkcija. Rezultatams įvertinti taikytos IoU, „F1“, preciziškumo ir specifiškumo metrikos. Kiekvienas modelis buvo apmokomas 8 valandas. 4.2.1 lentelėje pateikiami segmentavimo rezultatai. Vaizdiniai segmentavimo rezultatai parodyti A priede (1 pav.).

4.2.1 lentelė. Masačusetso kelių segmentavimo rezultatai, naudojant „Keras“ sistemą

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„EfficientNetB3“	62,1	73,5	64,1	75,9
FPN		60,5	72,5	61,1	76,6
„PSPNet“		58,6	70,9	59,1	74,9
„LinkNet“		61	72,7	62,9	75,6

Nagrinėjama atveju geriausią rezultatą parodė „U-Net“ modelis (62,1 % IoU). Pastebėta, kad visi keturi segmentavimo modeliai puikiai segmentuoja žiedines sankryžas. Manytina, kad rezultatai gali būti pagerinti, ilgiau apmokius tinklą ir naudojant „EfficientNetB7“ arba „SE-ResNetXt101“ enkoderius. Lyginant su [59] tyrėjų rezultatais, matyti, kad eksperimento metu buvo gauti prastesni rezultatai. Tai galėjo įvykti kelių priežasčių. Viena iš jų yra per trumpas mokymosi laikas. Kita priežastis yra skirtingų sistemų naudojimas. Vienos sistemos gali labiau optimizuoti skaičiavimus už kitas. Taip pat gali skirtis tikslo funkcijų ir įvertinimo metrikų skaičiavimo algoritmai.

Antruoju eksperimentu siekta:

- 1) remiantis [54], patikrinti, kokią įtaką rezultatams turi Tverskio židinio tikslo funkcija,
- 2) remiantis kitų tyrėjų darbais, patikrinti, kokią įtaką rezultatams turi aukštesnės raiškos vaizdai. Modelių įvestyje yra paduodami 512×512 dydžio vaizdai.

Tiesa, šiuo atveju buvo naudota „PyTorch“ sistema. Buvo lyginami „U-Net“, FPN ir „DeepLabV3+“ modeliai. Pabrėžiama, kad „U-Net“ modeliu buvo taikoma kombinuotoji tikslo funkcija, o pastarieji, naudojant Tverskio židinio funkciją. Kiekvienas modelis buvo apmokomas 8 valandas. Rezultatai pateikiami 4.2.2 lentelėje.

4.2.2 lentelė. Masačusetso kelių segmentavimo rezultatai naudojant „PyTorch“ sistemą

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„Xception“	59,74	72,35	75,02	75,3
FPN	„EfficientNetB3“	43,42	58,71	45,45	91,27
„DeepLabV3+“		39,13	54,48	40,82	91,26

Matyti, kad naudojant Tverskio židinio funkciją, gaunamas aukštas specifiškumo įvertis, bet žemas preciziškumo įvertis. Šiuo atveju segmentuoti keliai gauti platesni nei yra iš tikrųjų. Tai vaizduojama A priede (2 pav.). Aukštesnės raiškos vaizdai šiuo atveju didelės įtakos neturėjo. Manytina, kad tai galėjo nutikti, dėl skirtingų sistemų naudojimo.

Trečiuoju eksperimentu buvo siekta iširti, kokią įtaką rezultatams turi:

- 1) netiksliai sužymėtos kaukės,
- 2) mažas duomenų kiekis,
- 3) žemos erdvinės skyros vaizdai.

Šiems tikslams pasiekti buvo naudojamas specialiai šiam darbui sukurtas Lietuvos kelių duomenų rinkinys. Buvo atkartoti pirmojo eksperimento metu daryti žingsniai. Modelių įvestyje paduodamos 320×320 pikselių dydžio nuotraukos. Šį kartą nuspręsta taikyti tik „Dice“ tikslo funkciją. Kiekvienas modelis buvo apmokomas 6 valandas. Gauti rezultatai pateikti 4.2.3 lentelėje. Nagrinėjama atveju geriausius skaitinius rezultatus parodė FPN modelis (42 % IoU), tačiau iš A priedo (3 pav.) matyti, kad „U-Net“ modelio segmentavimo kaukė yra natūralesnė.

4.2.3 lentelė. Lietuvos kelių segmentavimo rezultatai, naudojant „Keras“ sistemą

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„EfficientNetB3“	41,8	57,3	58,3	59,1
FPN		42	57,4	56,9	60,8
„PSPNet“		41	56,5	54,9	59,8
„LinkNet“		41,6	57,2	56,6	60,9

Darytina išvada, kad turint blogai sužymėtas kaukes, giliajam dirbtinių neuronų tinklui yra labai sunku išmokti tam tikrų požymių, kadangi vienu ar kitu atveju tie požymiai bus klaidingi (pvz., šalikelės sužymėtos kaip kelias). Pagaliau tinklas nepasieks aukštų rezultatų. Taigi, atliekant tyrimus yra labai svarbu atrinkti duomenis, kurie turi kokybiškai sužymėtas kaukes.

Ketvirtuoju eksperimentu siekta patikrinti kokią įtaką rezultatams turi skirtingi enkoderiai. Remiantis [54, 76, 77], buvo pasirinkti „EfficientNetB3“, „MobileNetV3L“ ir „SE-ResNetXt101“ enkoderiai. „MobileNet“ tipo enkoderiai buvo specialiai sukurti mobiliesiems įrenginiams. Jie pasižymi mažu parametrų skaičiumi, todėl greičiau mokosi. Be to, jų rezultatai, sprendžiant vaizdų klasifikavimo uždavinį nežymiai skiriasi nuo kitų daug parametrų turinčių enkoderių. „SE-ResNetX101“ enkoderis, atvirkščiai, turi labai daug parametrų, todėl jį apmokyti, reikia daug išteklių ir laiko. „EfficientNetB3“ yra vienas iš moderniausių požymių išskyrimo tinklų, pasižymintis tiek sparčiu mokymusi, tiek aukštu tikslumu. Šį kartą buvo naudotas „DeepGlobe“ duomenų rinkinys. Naudojant „PyTorch“ sistemą buvo patikrinta daugiau įvairių modelių. Kiekvienas modelis buvo apmokomas 8 val. „DeepLabV3+“ ir „SE-ResNetXt101“ konfigūracijai buvo taikoma Tverskio židinio funkcija, o visoms kitoms konfigūracijoms – „Dice“ tikslo funkcija. Rezultatai pateikti 4.2.4 lentelėje. Lyginamųjų modelių prognozuotos segmentavimo kaukės parodomos A priede (4 pav.).

4.2.4 lentelė. „DeepGlobe“ kelių segmentavimo rezultatai, naudojant „PyTorch“ sistemą

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„MobileNetV3L“	57,4	68,08	72,32	76,56
„U-Net++“	„EfficientNetB3“	58,95	69,55	69,26	82,67
FPN	„EfficientNetB3“	58,95	69,55	69,26	82,67
„DeepLabV3+“	„MobileNetV3L“	59,37	70,02	71,65	79,82
	„EfficientNetB6“	57,4	68,08	72,32	76,56
	„SE-ResNetXt101“	32,1	44,29	32,99	95,53
„MANet“	„EfficientNetB3“	57,86	68,41	71,7	78,27

Nagrinėjamu atveju, geriausius rezultatus parodė „DeepLabV3+“ modelis, turintis „MobileNetV3L“ enkoderį (59,37 % IoU). Gauti rezultatai grindžiami tuo, kad per tą patį laiką „MobileNetV3L“

enkoderį turintys segmentavimo modeliai, mokėsi žymiai greičiau, t. y. buvo atliktas didesnis epochų skaičius. Manytina, kad pailginus mokymo trukmę, „EfficientNetB6“ enkoderį turintis modelis galėtų pasiekti geriausią rezultatą. Tai pasitvirtina 4.2.2 skirsnyje, atlikus pirmąjį pastatų segmentavimo eksperimentą.

4.2.2. Pastatų segmentavimas

Pirmasis pastatų segmentavimo eksperimentas atliktas, naudojant Masačusetso pastatų duomenų rinkinį ir „DeepLabV3+“ segmentavimo modelį. Modelio įvestyje buvo naudojami 384×384 pikselių dydžio vaizdai. Buvo pasirinkti „MobileNetV3S“ ir „SE-ResNetXt50“ enkoderiai. Buvo taikoma „Dice“ tikslo funkcija. Po 4 valandų mokymo, geriausias rezultatas buvo gautas naudojant „SE-ResNetXt50“ enkoderį. Modelis per tokį laiką pasiekė 67,47 % IoU įvertį (57,94 % IoU, naudojant „MobileNetV3S“). Šis eksperimentas geriau parodė skirtingų enkoderių įtaka segmentavimo modeliui. Vaizdiniai segmentavimo rezultatai pateikti A priede (5 pav.).

Antrojo eksperimento metu, remiantis [63], buvo atliktas išsamesnis pastatų segmentavimo tyrimas. Buvo naudotas „INRIA“ pastatų duomenų rinkinys ir „PyTorch“ sistema. Lyginamųjų modelių prognozuotos segmentavimo kaukės parodytos A priede (6 pav.).

4.2.5 lentelė. „INRIA“ pastatų segmentavimo rezultatai, naudojant „PyTorch“ sistemą

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„MobileNetV3L“	67,82	77,43	71,46	92,71
„U-Net++“	„EfficientNetB3“	66,99	76,51	70,52	92,67
FPN	„EfficientNetB3“	75,47	83,33	86,9	84,81
	„SE-ResNetXt101“	68,96	78,27	72,5	92,84
„DeepLabV3+“	„SE-ResNetXt101“	75,79	83,93	86,25	85,48
	„MobileNetV3S“	70,13	79,09	82,39	82,4
	„MobileNetV3L“	73,55	82,11	85,55	83,31
„MANet“	„EfficientNetB3“	69,59	78,81	73,2	92,61

Nagrinėjama atveju geriausius rezultatus parodė „DeepLabV3+“ ir FPN modeliai (75,79 % IoU ir 75,47 % IoU atitinkamai). Iš esmės rezultatai panašūs į [63] rezultatus. Matyti, kad visi modeliai prastai išlaiko tikrąją pastatų formą, tačiau apima didžiąją dalį pastatų. Taip pat pastebima, kad segmentavimo modeliams sunku atskirti tankiai išdėstytus pastatus ir juos kartais apjungia. Svarbu paminėti, kad viena epocha, naudojant „DeepLabV3+“ ir „SE-ResNetXt101“ konfigūraciją, truko penkis kartus ilgiau, nei naudojant „DeepLabV3+“ ir „MobileNetV3L“ konfigūraciją, o rezultatai skiriasi tik dviem procentiniais punktais.

4.3. Daugelio klasių segmentavimo rezultatai

Daugelio klasių segmentavimo eksperimentams atlikti buvo pasirinktas „DroneDeploy“ duomenų rinkinys. Naudojant „PyTorch“ sistemą ir [74], buvo lyginami „U-Net“, „U-Net++“, FPN, „PSPNet“, „PAN“, „MANet“ ir „DeepLabV3+“ segmentavimo modeliai. Buvo pasirinkta „Dice“ tikslo funkcija. Rezultatai įvertinti taikant IoU, „F1“, preciziškumo ir specifiškumo metrikos.

Modelis	Enkoderis	IoU, %	„F1“, %	Preciziškumas, %	Specifiškumas, %
„U-Net“	„EfficientNetB3“	75,45	82,99	93,05	82,93
„U-Net++“		82,02	88,59	88,60	88,58
FPN		80,48	87,19	87,20	87,17
„PSPNet“		73,36	81,44	81,46	81,42
„PAN“		77,23	85,06	85,26	84,87
„MANet“		78,58	85,70	85,78	85,62
„DeepLabV3+“	„MobileNetV3L“	78,73	85,88	85,92	85,83
	„EfficientNetB3“	79,44	86,41	86,49	86,34
	„SE-ResNetXt101“	78,03	85,40	85,49	85,31

Matyti, kad geriausią skaitinį ir vaizdinį rezultatą parodė „U-Net++“ segmentavimo modelis (82,02 % IoU). Lyginant su [5, 64, 66] rezultatais matyti, kad buvo gauti gerokai geresni rezultatai. Taip išėjo, nes kiekvienai klasei skaičiuojant IoU buvo įtrauktas fonas, kuris sudaro didžiąją dalį vaizdo. Taigi, šis eksperimentas taip pat parodė, kad reikia būtinai atkreipti dėmesį į tai, kaip yra skaičiuojamas modelio tikslumas. A priede (9 pav.) pateikti lyginamųjų modelių vaizdiniai rezultatai. Matyti, kad skiriasi kai kurių modelių prognozuotų segmentų spalvos. Taip gavosi, nes modelių išvestyje buvo pateiktas vadinamasis karščio žemėlapis (angl. *Heatmap*), kuris tiesiog užkoduoja skirtingas klases taip, kad matytųsi aiškus skirtumas. Pateiktame pavyzdyje, vieni modeliai išskyrė daugiau klasių, todėl pasikeitė grafinis rezultatų pateikimas. Taip pat reikia pabrėžti, kad yra galimi du daugelio klasių segmentavimo atvejai. Pirmuoju atveju, kiekviena klasė gali būti priskiriama koduojant ją tam tikra intensyvumo reikšme, pvz. 0 – fonas, 1 – pastatai ir t. t. Antruoju atveju, kiekviena klasė gali būti priskiriama koduojant ją kaip atskirą kanalą. Vėl gi, nuo to priklausys galutiniai segmentavimo rezultatai.

4.4. Trumpas skyriaus apibendrinimas

Apibendrinant kelių segmentavimo eksperimentus galima sakyti, kad giliojo mokymosi metodai neblogai išskiria aiškiai matomus ilgus kelius, ir blogiau išskiria siaurus ir trumpus kelius. Žemas IoU rodiklis nebūtinai rodo prastą galutinį rezultatą todėl, kad tam tikrais atvejais, segmentavimo modeliai geba išskirti realiai egzistuojančius kelius, kurie nebuvo sužymėti ekspertų. Pavyzdžiui, A priede (žr.

1 pav.) matyti, kad originalioje kaukėje nebuvo sužymėta žiedinė sankryža, tačiau visi modeliai sugebėjo ją teisingai segmentuoti. Darytina išvada, kad kuo tiksliau bus sužymėti duomenys, tuo tiksliau modelis galės juos segmentuoti. Segmentavimo rezultatams didelę įtaką turi tinkamai pasirinkta tikslo funkcija. Galima sakyti, kad nagrinėjant asfaltinius kelius, geriausius skaitinius ir vaizdinius rezultatus rodo „U-Net“ modelis, o nagrinėjant įvairesnio tipo kelius, labiau tinka „DeepLabV3+“ modelis. Apibendrinant pastatų segmentavimo rezultatus galima teigti, kad giliojo mokymosi metodai geba išskirti įvairaus dydžio pastatus nuo fono, tačiau sunkiau susitvarko išlaikydami tikrąją jų formą. Remiantis eksperimentais ir [63, 64], galima sakyti, kad papildomų blokų, tokių kaip KEP ar IELPS taikymas gerina galutinius rezultatus, kadangi yra atsižvelgiama į kontekstinę informaciją (FPN, „DeepLabV3+“). Taip pat darytina išvada, kad aukštesnės raiškos ir erdvinės skyros vaizdai geriau atskiria tikrąsias klases, todėl prognozuotos kaukės yra tikslesnės. Apibendrinant daugelio klasių segmentavimo rezultatus galima sakyti, kad tai yra viena iš sudėtingiausių užduočių kompiuterinės regos srityje. Prasti rezultatai gali būti siejami su tuo, kad šiuo metu vis dar yra per mažai duomenų rinkinių, skirtų daugelio klasių segmentavimo uždaviniams spręsti. Nors dabartiniai giliojo mokymosi metodai nepasižymi aukštu daugelio klasių segmentavimo tikslumu, tačiau jie žymiai geriau išskiria įvairioms klasėms priklausančius objektus už klasikinius, pavyzdžiui, klasterizavimo metodus. Be to, yra smarkiai palengvinama duomenų žymėjimo užduotis.

APIBENDRINIMAS. IŠVADOS

Šiame darbe buvo ištirtos giliojo mokymosi metodų, taikytinų vaizdams segmentuoti, galimybės. Pirma, buvo susipažinta su giliaisiais dirbtinių neuronų tinklais. Išnagrinėti įvairūs vaizdų segmentavimo metodai. Pateiktos pamatinės sąsūkos neuroninių tinklų architektūros. Antra, buvo aptarti pagrindiniai aerofotografinių nuotraukų analizės iššūkiai. Analizuojami modernūs segmentavimo modeliai, taikytini aerofotografinėms nuotraukoms analizuoti. Nagrinėtos segmentavimui skirtos tikslo funkcijos ir segmentavimo rezultatų įvertinimo metrikos. Atliktas dvejetainio ir daugelio klasių segmentavimo tyrimas analizuojant kitų tyrėjų darbus. Trečia, buvo išrinkti ir apžvelgti viešai prieinami aerofotografinių nuotraukų rinkiniai. Pagaliau naudojant nemokamas debesijos paslaugas, buvo atlikti kelių, pastatų ir daugelio klasių segmentavimo eksperimentai.

Atlikus išsamius tyrimus ir papildomus eksperimentus darytina išvada, kad giliojo mokymosi metodai yra labai perspektyvūs, analizuojant aerofotografines nuotraukas. SNT pranoksta visus kitus klasikinius segmentavimo metodus. Pirma, giliaji dirbtinių neuronų tinklai neblogai segmentuoja aiškiai matomus ilgus kelius, bet blogiau išskiria siaurus ir trumpus kelius. Iš nagrinėtų modelių galima išskirti „PP-LinkNet“ modelį, kuris pasiekė 70,53 % IoU vertę, naudojant „DeepGlobe“ duomenų rinkinį (4 % pranašesnis už „DeepGlobe“ varžybų nugalėtoją). Šis modelis geriau nei kiti išskiria siauresnius ir trumpesnius kelius. Be to, pastebima mažiau trūkių. „PP-LinkNet“ modelis parodė geras apibendrinimo galimybes. Antra, giliojo mokymosi metodai neblogai geba išskirti įvairaus dydžio pastatus ir kitus objektus, tačiau sunkiau susitvarko išlaikydami tikrąją objektų formą. Iš nagrinėtų segmentavimo modelių galima išskirti modernųjį „PFNet“. Šis modelis, lyginant su kitais, pasiekia geriausią tikslumo ir skaičiavimo sąnaudų santykį. Naudojant Faihingo prie Enco ir Potsdamo duomenų rinkinius, buvo pasiekti 70,4 % IoU ir 75,4 % IoU įverčiai (iki 5 % geresni rezultatai, lyginant su kitais modeliais). „PFNet“ prognozuotos kaukės atrodo natūraliausiai, nes yra išlaikomi aštrūs pastatų kampai. Kiti modeliai yra linkę užapvalinti kampus. Pagaliau giliojo mokymosi metodai gali padėti spręsti daugelio klasių segmentavimo problemas. Šiuo atveju taip pat galima išskirti „PFNet“ modelį. Pavyzdžiui, naudojant „iSAID“ duomenų rinkinį, „PFNet“ yra 5 % tikslesnis pagal mIoU įvertį, lyginant su kitais modeliais. Dėl aerofotografinių nuotraukų sudėtingumo ir objektų įvairovės susiduriama su daugybe iššūkių, įskaitant daugiamatiškumą, neapibrėžtumą, netiesiškumą ir kt., todėl dabartiniais segmentavimo metodams trūksta tikslumo, vis dar reikia naudoti žmogaus intelektą. Kita vertus, giliojo mokymosi metodai stipriai sumažina duomenų žymėjimo proceso trukmę.

Giliųjų neuroninių tinklų rezultatai priklauso nuo duomenų kiekio, duomenų kaukių kokybės, vaizdų raiškos, vaizdų erdvinės skyros, tikslo funkcijų, mokymo trukmės ir kt. Dėl pasiektos

pažangos kompiuterinės regos ir jutiklinių technologijų srityse, nuotolinio aptikimo vaizdų segmentavimo rezultatai per pastaruosius metus labai pagerėjo. Nuotolinio aptikimo sritis tik dabar įgauna pagreitį, tyrėjai kuria ir dalinasi vis kokybiškesniais duomenų rinkiniais ir tikslesniais segmentavimo modeliais, todėl kuo toliau, tuo vaizdų analizės procesas reikalaus mažiau žmogaus įsikišimo.

Taigi iškeltas darbo tikslas buvo pasiektas, o darbo uždaviniai nuosekliai įvykdyti: atlikta automatinės aerofotografinių nuotraukų analizės metodų analitinė apžvalga; atrinkti giliojo mokymosi metodai, perspektyvius taikyti automatinei aerofotografinių nuotraukų analizei. Įgyvendinti ir eksperimentiškai ištirti atrinkti automatinės aerofotografinių nuotraukų analizės metodai ir pateikti tyrimo rezultatai.

LITERATŪRA

1. Yuheng, S. & Yan H. (2017). *Image Segmentation Algorithms Overview*. Prieiga per internetą: <https://arxiv.org/ftp/arxiv/papers/1707/1707.02051.pdf>
2. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). *Image Segmentation Using Deep Learning: A Survey*. <http://doi.org/10.1109/TPAMI.2021.3059968>
3. Bhattacharjee, G., & Pujari, S. K. (2017). Aerial Image Segmentation: A Survey. *International Journal of Applied Information Systems (IJ AIS)* – ISSN : 2249-0868. <http://doi.org/10.5120/ijais2017451702>
4. Sisodiya N., Dube, N., & Thakkar, P. (2020) Next-Generation Artificial Intelligence Techniques for Satellite Data Processing. *Artificial Intelligence Techniques for Satellite Image Analysis*. https://doi.org/10.1007/978-3-030-24178-0_11
5. Demir, I., Koperski, K., Lindenbaum, D., et al. (2018). DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 172-17209. <https://doi.org/10.1109/CVPRW.2018.00031>
6. Li, Z., Yang, W., Peng, S. & Liu, F. (2020). *A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects*. Prieiga per internetą: <https://arxiv.org/ftp/arxiv/papers/2004/2004.02806.pdf>
7. Shaurya, M. (2019). *McCulloch-Pitts Neuron vs Perceptron model*. Prieiga per internetą: <https://medium.com/@manushaurya/mcculloch-pitts-neuron-vs-perceptron-model-8668ed82c36>
8. Chandra, A. L. (2018). *McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron*. Prieiga per internetą: <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>
9. *Perceptron*. (2020). Wikipedia. Prieiga per internetą: <https://nl.wikipedia.org/wiki/Perceptron>
10. Stutz, D. (2014). *Recognizing Handwritten Digits using a Two-Layer Perceptron and the MNIST Dataset*. Prieiga per internetą: <https://davidstutz.de/recognizing-handwritten-digits-mnist-dataset-twolayer-perceptron/>
11. Bengio, Y. (2009). *Learning Deep Architectures for AI*. Prieiga per internetą: https://www.researchgate.net/publication/215991023_Learning_Deep_Architectures_for_AI
12. Biswal, A. (2020). *Top 10 Deep Learning Algorithms You Should Know in 2021*. Prieiga per internetą: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm>
13. Ankile, L. L., Heggland, M. F. & Krange, K. (2020). *Deep Convolutional Neural Networks: A survey of the foundations, selected improvements, and some current applications*. Prieiga per internetą: <https://arxiv.org/abs/2011.12960>
14. Khan, A., Sohail, A., Zahoora, U. et al. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev* 53, 5455–5516 <https://doi.org/10.1007/s10462-020-09825-6>
15. Sumit, S. (2018). *A Comprehensive Guide to Convolutional Neural Networks – the ELI5 way*. Prieiga per internetą: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

16. Aloysius, N. M., Geetha, M. (2017). A review on deep convolutional neural networks. *2017 International Conference on Communication and Signal Processing (ICCSP)*: 0588–0592.
<https://doi.org/10.1109/ICCSP.2017.8286426>
17. Dirbtinis neuroninis tinklas. (2021). Wikipedia. Prieiga per internetą:
https://lt.wikipedia.org/wiki/Dirbtinis_neuroninis_tinklas
18. Krishna, K. (2018). *Introduction to Exponential Linear Unit*. Prieiga per internetą:
<https://medium.com/@krishnakalyan3/introduction-to-exponential-linear-unit-d3e2904b366c>
19. Pavlovsky, V. (2017). *Introduction To Convolutional Neural Networks*. Prieiga per internetą:
<https://www.vojtech.net/posts/intro-convolutional-neural-networks/>
20. *Convolutional Neural Networks (CNN): Step 3 – Flattening*. (2018) . Super Data Science. Prieiga per internetą:
<https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening>
21. *Fully-Connected Layer with dynamic input shape*. (2019). Medium. Prieiga per internetą:
<https://medium.com/@tecokids.monastir/fully-connected-layer-with-dynamic-input-shape-70c869ae71af>
22. Gu, J., Wang, Z., Kuen, J. et al. (2015). *Recent Advances in Convolutional Neural Networks*. Prieiga per internetą: <https://arxiv.org/pdf/1512.07108.pdf>
23. Chablani, M. (2017). *Deep learning concepts – PART 1*. Prieiga per internetą:
<https://towardsdatascience.com/deep-learning-concepts-part-1-ea0b14b234c8>
24. Wu, G., Shao, X., Guo, Z. et al. (2018). Automatic Building Segmentation of Aerial Imagery Using Multi-Constraint Fully Convolutional Networks. *Remote Sens.* 10, 407. <https://doi.org/10.3390/rs10030407>
25. Kaur, D. & Kaur, Y. (2014). *Various Image Segmentation Techniques: A Review*. Prieiga per internetą:
<https://ijcsmc.com/docs/papers/May2014/V3I5201499a84.pdf>
26. Jayapriya, P. & Hemalatha, S. (2019). *Comparative Analysis Of Image Segmentation Techniques And Its Algorithm*. Prieiga per internetą: <http://www.ijstr.org/final-print/oct2019/Comparative-Analysis-Of-Image-Segmentation-Techniques-And-Its-Algorithm.pdf>
27. Bhargavi, K. & Jyothi, S. (2014). *A Survey on Threshold Based Segmentation Technique in Image Processing*. Prieiga per internetą:
https://www.researchgate.net/publication/309209325_A_Survey_on_Threshold_Based_Segmentation_Technique_in_Image_Processing
28. Albregtsen, F. (1993). *Non-Parametric Histogram Thresholding Methods - Error Versus Relative Object Area*. Prieiga per internetą: https://www.researchgate.net/publication/228567852_Non-parametric_histogram_thresholding_methods-error_versus_relative_object_area
29. Bedros, S. (2017). *Edge Detection*. Prieiga per internetą:
<http://dept.me.umn.edu/courses/me5286/vision/VisionNotes/2017/ME5286-Lecture7-2017-EdgeDetection2.pdf>
30. Wirth, M. (2004). *Image Segmentation: Edge-based*. Prieiga per internetą:
<http://www.cyto.purdue.edu/cdroms/micro2/content/education/wirth05.pdf>

31. The University of Auckland. *Edge Detection*. Prieiga per internetą: https://www.cs.auckland.ac.nz/compsci373s1c/PatricesLectures/Edge%20detection-Sobel_2up.pdf
32. Kaur, M. & Goyal, M. (2013). *A Review on Region Based Segmentation*. Prieiga per internetą: <https://ijsr.net/archive/v4i4/27041504.pdf>
33. Meyer, F. & Beucher, S. (1990). Morphological Segmentation. In *Journal of Visual Communication and Image representation*, vol. 1, no. 1, pp. 21-46. [https://doi.org/10.1016/1047-3203\(90\)90014-M](https://doi.org/10.1016/1047-3203(90)90014-M)
34. Beucher, S. (1992). *THE WATERSHED TRANSFORMATION APPLIED TO IMAGE SEGMENTATION*. Prieiga per internetą: <https://folk.uib.no/eha070/mat262/papers/Beucher.pdf>
35. Viswarupan, N. (2017). *K-Means Data Clustering*. Prieiga per internetą: <https://towardsdatascience.com/k-means-data-clustering-bce3335d2203>
36. Hassan, M. (2018). *VGG16 – Convolutional Network for Classification and Detection*. Prieiga per internetą: <https://neurohive.io/en/popular-networks/vgg16/>
37. He, K., Zhang, X., Ren, S. & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. Prieiga per internetą: <https://arxiv.org/pdf/1512.03385.pdf>
38. Jeong, Y. W. (2020). *ResNet: The Most Popular Network in the Computer-Vision Era*. Prieiga per internetą: <https://towardsdatascience.com/resnet-the-most-popular-network-in-computer-vision-era-973df3e92809>
39. Szegedy, C., Liu, W., Jis, Y. et al. (2014). *Going Deeper with Convolutions*. Prieiga per internetą: <https://arxiv.org/pdf/1409.4842.pdf>
40. Anwar, A. (2019). *Difference between AlexNet, VGGNet, ResNet, and Inception*. Prieiga per internetą: <https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaecccc96>
41. Patel, K. (2020). *Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet*. Prieiga per internetą: <https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inception-densenet-beb8b116866d>
42. Huang, G., Liu, Z., Maaten, L. & Weinberger, K. Q. (2016). *Densely Connected Convolutional Networks*. Prieiga per internetą: <https://arxiv.org/pdf/1608.06993.pdf>
43. Xing, Y., Zhong, L. & Zhong, X. (2020). An Encoder-Decoder Network Based FCN Architecture for Semantic Segmentation. *Wireless Communications and Mobile Computing*, vol. 2020. <https://doi.org/10.1155/2020/8861886>
44. Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G. (2020). Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3735-3756. <http://doi.org/10.1109/JSTARS.2020.3005403>
45. Mehta, S. (2020). *Encoder-Decoder Networks for Semantic Segmentation*. Prieiga per internetą: https://courses.cs.washington.edu/courses/cse576/17sp/notes/Sachin_Talk.pdf
46. Parmar, V., Bhatia, N., Negi, S. & Suri, M. (2020). *Exploration of Optimized Semantic Segmentation Architectures for edge-Deployment on Drones*. Prieiga per internetą: <https://arxiv.org/pdf/2007.02839.pdf>

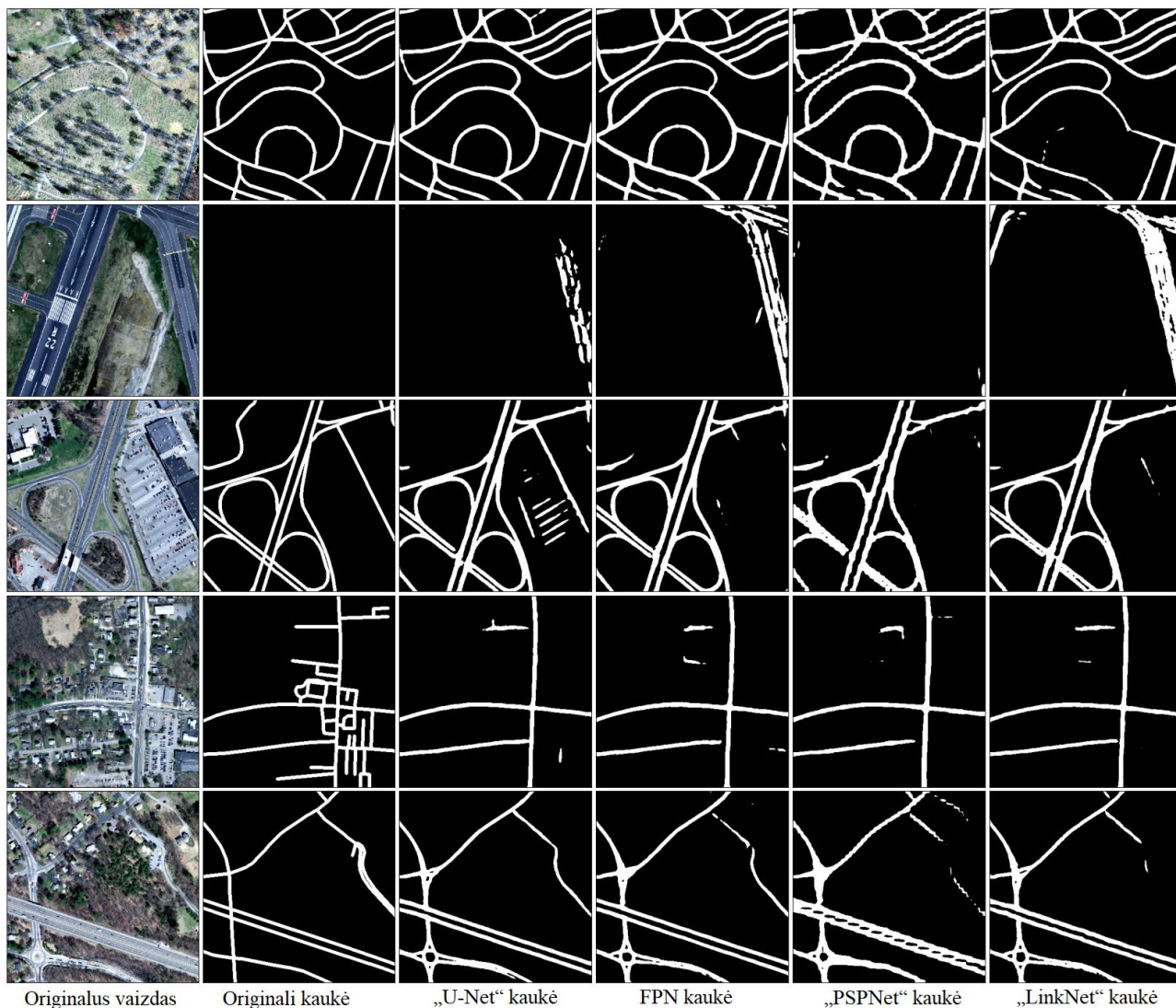
47. Long, J., Shelhamer, E. & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440. <http://doi.org/10.1109/CVPR.2015.7298965>
48. Ronneberger, O., Fischer, P. & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Prieiga per internetą: <https://arxiv.org/pdf/1505.04597.pdf>
49. Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B. & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936-944. <http://doi.org/10.1109/CVPR.2017.106>
50. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6230-6239. <http://doi.org/10.1109/CVPR.2017.660>
51. Chaurasia, A. & Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1-4. <http://doi.org/10.1109/VCIP.2017.8305148>
52. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. (2016). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-84. <http://doi.org/10.1109/TPAMI.2017.2699184>.
53. Chen, L., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Prieiga per internetą: <https://arxiv.org/pdf/1802.02611.pdf>
54. Jadon, S. (2020). A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1-7. <http://doi.org/10.1109/CIBCB48159.2020.9277638>
55. Cira, C., Alcarria, R., Manso-Callejo, M. & Serradilla F. (2020). A Deep Learning-Based Solution for Large-Scale Extraction of the Secondary Road Network from High-Resolution Aerial Orthoimagery. *Applied Sciences*. 10(20):7272. <https://doi.org/10.3390/app10207272>
56. Wulamu, A., Shi, Z., Zhang, D. & He, Z. (2019). Multiscale Road Extraction in Remote Sensing Images. *Computational Intelligence and Neuroscience*, vol. 2019. <https://doi.org/10.1155/2019/2373798>
57. Berman, M., Triki, A. & Blaschko, M. (2018). *The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks*. Prieiga per internetą: <https://arxiv.org/pdf/1705.08790.pdf>
58. Tran, A., Zonoozi, A., Varadarajan, J. & Kruppa, H. (2020). PP-LinkNet: Improving Semantic Segmentation of High Resolution Satellite Imagery with Multi-stage Training. *In Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC'20)*. <https://doi.org/10.1145/3423323.3423407>

59. Abderrahim, N., Abderrahim, S. & Rida, A. (2020). Road Segmentation using U-Net architecture. 2020 *IEEE International conference of Moroccan Geomatics (Morgeo)*, pp. 1-4. <http://doi.org/10.1109/Morgeo49228.2020.9121887>
60. Abdollahi, A., Pradhan, B. & Alamri, A. (2020). VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data. *IEEE Access*, vol. 8, pp. 179424-179436. <http://doi.org/10.1109/ACCESS.2020.3026658>
61. Sebastian, C., Imbriaco, R., Bondarev, E. & de With, P. (2020). *Adversarial Loss for Semantic Segmentation of Aerial Imagery*. Prieiga per internetą: <https://arxiv.org/pdf/2001.04269.pdf>
62. Ozaki, K. (2017). *Winning Solution for the Spacenet Challenge: Joint Learning with OpenStreetMap*. Prieiga per internetą: <https://i.ho.lc/winning-solution-for-the-spacenet-challenge-joint-learning-with-openstreetmap.html>
63. Sebastian, C., Imbriaco, R., Bondarev, E. & de With, P. (2020). *Contextual Pyramid Attention Network for Building Segmentation in Aerial Imagery*. Prieiga per internetą: <https://arxiv.org/pdf/2004.07018.pdf>
64. Li, X., He, H., Li, X. et al. (2021). *PointFlow: Flowing Semantics Through Points for Aerial Image Segmentation*. Prieiga per internetą: <https://arxiv.org/pdf/2103.06564.pdf>
65. Azimi, S. M., Henry, C., Sommer, L., Schumann, A., & Vig, E. (2019). SkyScapes - Fine-Grained Semantic Understanding of Aerial Scenes. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7392-7402. <http://doi.org/10.1109/ICCV.2019.00749>
66. Heffels., M. R. & Vamschoren, J. (2020). *Aerial Imagery Pixel-level Segmentation*. Prieiga per internetą: <https://arxiv.org/pdf/2012.02024.pdf>
67. Kumar, S. (2020). *Data splitting technique to fit any Machine Learning Model*. Prieiga per internetą: <https://towardsdatascience.com/data-splitting-technique-to-fit-any-machine-learning-model-c0d7f3f1c790>
68. Mnih, V. (2013). *Machine Learning for Aerial Image Labeling*. Prieiga per internetą: https://www.cs.toronto.edu/~vmnih/docs/Mnih_Volodymyr_PhD_Thesis.pdf
69. Maggiori, E., Tarabalka, Y., Charpiat, G. & Alliez, P. (2017). Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark. 2017 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. <http://doi.org/10.1109/IGARSS.2017.8127684>
70. Pilkington, N., Svetlichnaya, S. & Holmes, T. (2019). *DroneDeploy Machine Learning Segmentation Benchmark*. Prieiga per internetą: <https://github.com/dronedeploy/dd-ml-segmentation-benchmark>
71. Ma, R., Tao, P. & Tang, H. (2019). Optimizing Data Augmentation for Semantic Segmentation on Small-Scale Dataset. *Proceedings of the 2nd International Conference on Control and Computer Vision (ICCCV 2019)*. <https://doi.org/10.1145/3341016.3341020>
72. Kumar, H. (2021). Data augmentation Techniques. Prieiga per internetą: <https://iq.opengenus.org/data-augmentation/>
73. Yakubovskiy, P. (2020). *Python library with Neural Networks for Image Segmentation based on Keras and TensorFlow*. Prieiga per internetą: https://github.com/qubvel/segmentation_models

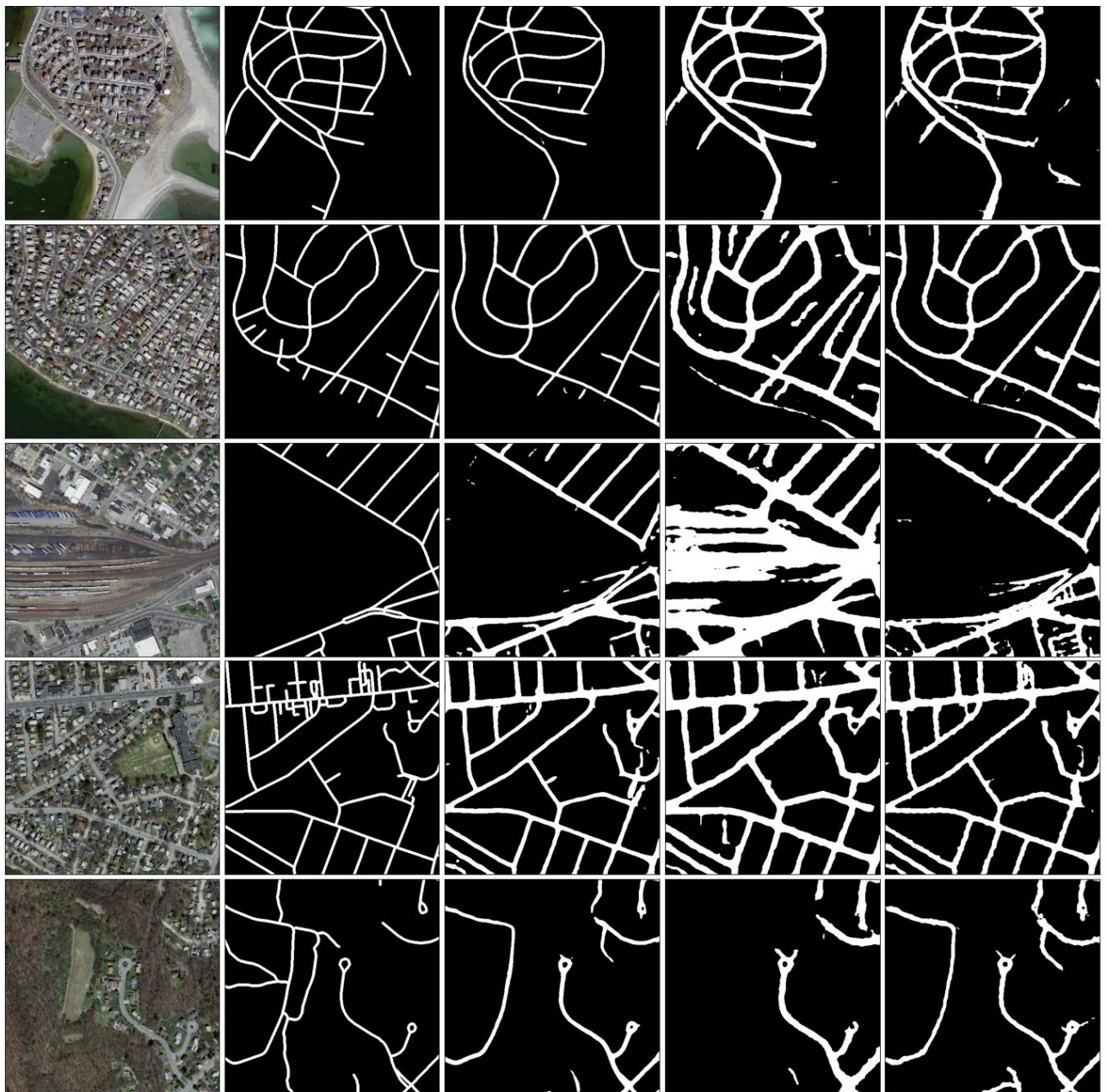
74. Yakubovskiy, P. (2020). *Python library with Neural Networks for Image Segmentation based on PyTorch*. Prieiga per internetą: https://github.com/qubvel/segmentation_models.pytorch
75. Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M. & Kalinin, A.A. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information* 2020, 11, 125. <https://doi.org/10.3390/info11020125>
76. Tan, M. & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Prieiga per internetą: <https://arxiv.org/pdf/1905.11946.pdf>
77. Howard, A., Sandler, M., Chu, G. et al. (2019). *Searching for MobileNetV3*. Prieiga per internetą: <https://arxiv.org/pdf/1905.02244.pdf>

PRIEDAI

A priedas. Automatinės aerofotografinių nuotraukų analizės metodų eksperimentinio tyrimo rezultatai



Šiame paveiksle pateikti Masačusetso kelių duomenų rinkinio segmentavimo pavyzdžiai.



Originalus vaizdas

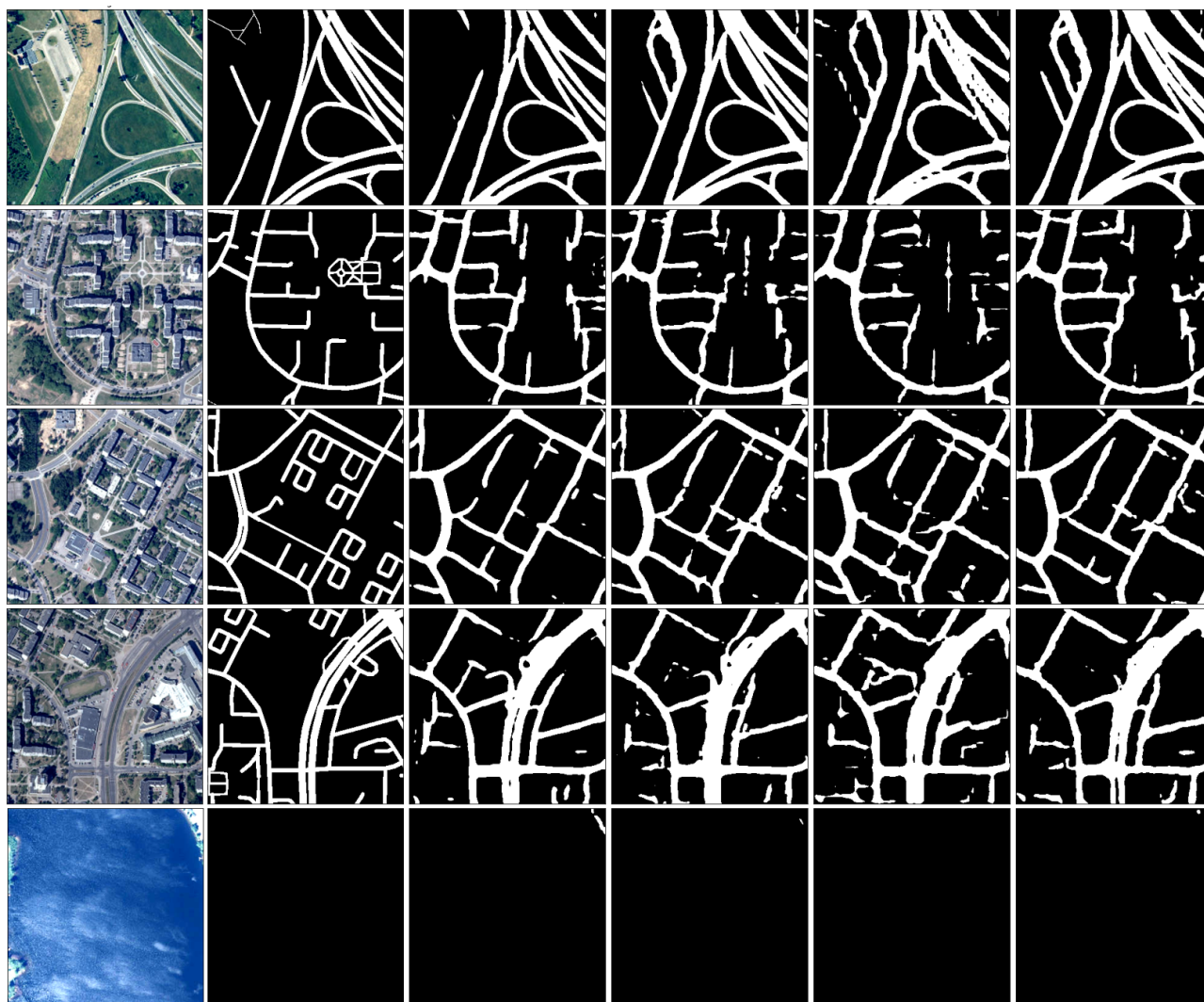
Originali kaukė

„U-Net“ kaukė

FPN kaukė

„DeepLabV3+“ kaukė

Šiame paveiksle pateikti Masačusetso kelių duomenų rinkinio segmentavimo pavyzdžiai. Atkreipiamas dėmesys, kad taikant Tverskio židinio tikslo funkciją, keliai tampa platesni nei yra iš tikrųjų. Be to, PPT modelis geležinkelio bėgius klasifikavo kaip kelius.



Originalus vaizdas

Originali kaukė

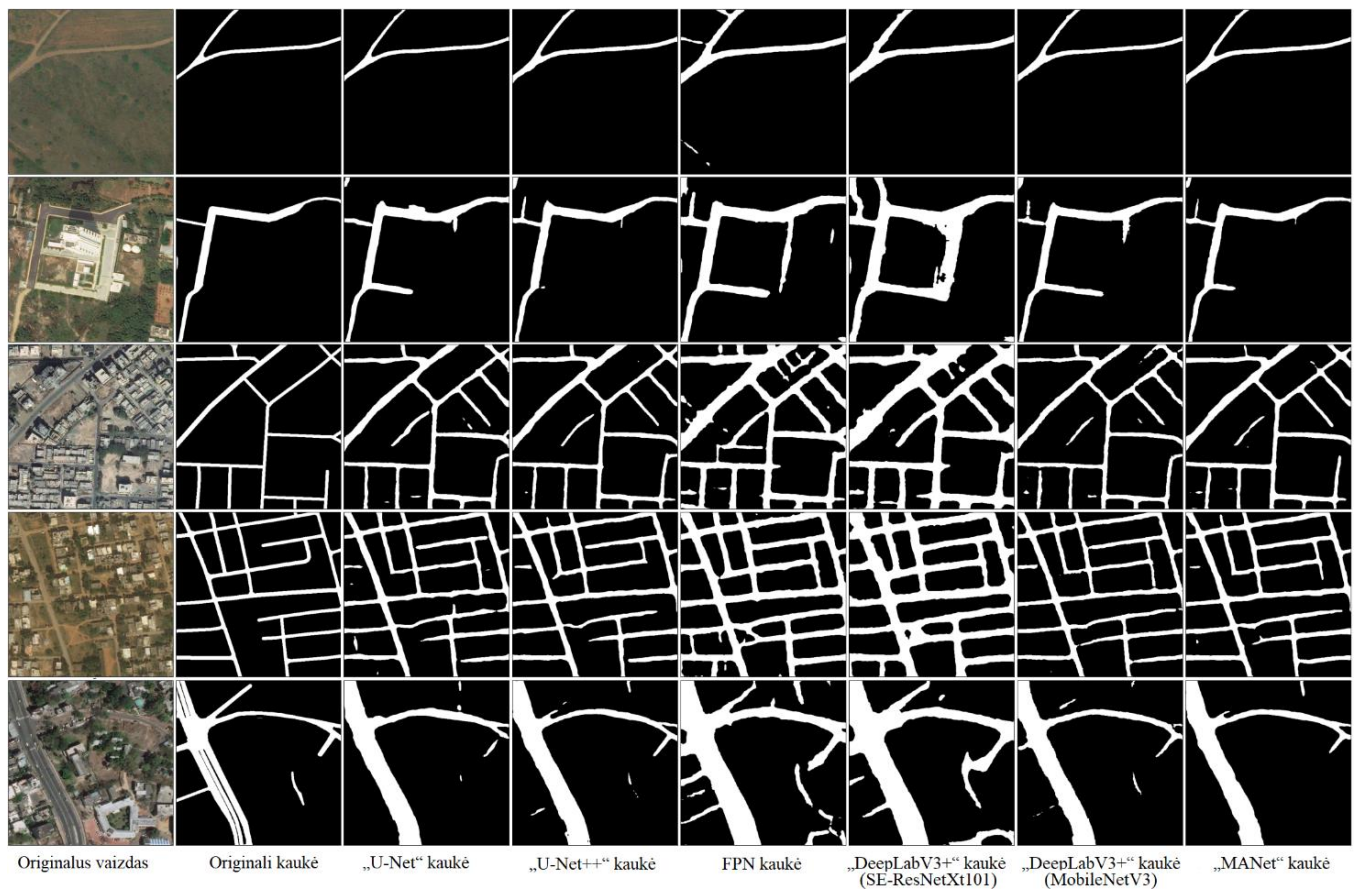
„U-Net“ kaukė

FPN kaukė

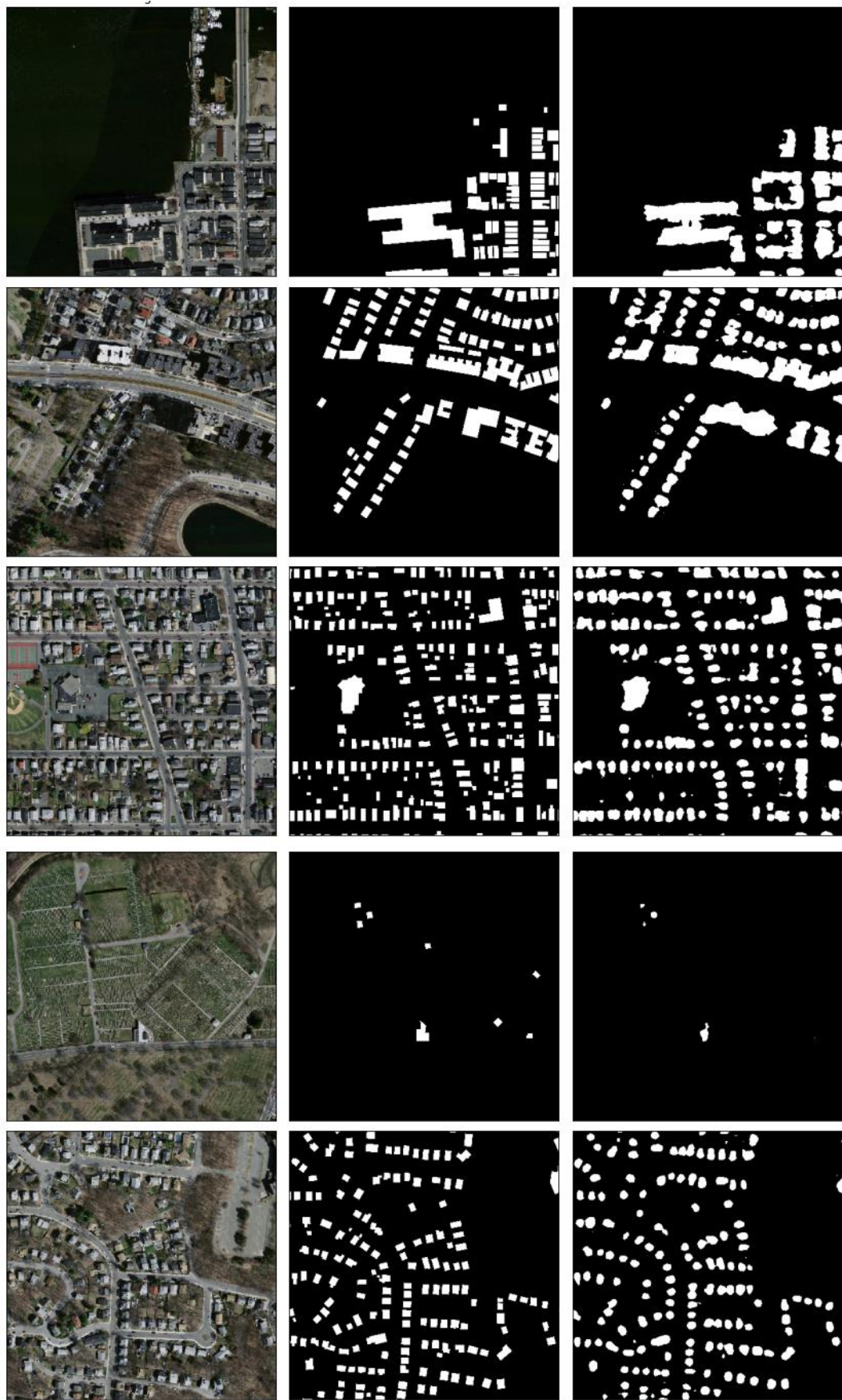
„PSPNet“ kaukė

„LinkNet“ kaukė

Šiame paveiksle pateikti Lietuvos kelių segmentavimo pavyzdžiai.



Šiame paveiksle pateikti „DeepGlobe“ duomenų rinkinio segmentavimo pavyzdžiai. Vėl gi, platesni keliai buvo gauti taikant Tverskio židinio funkciją.

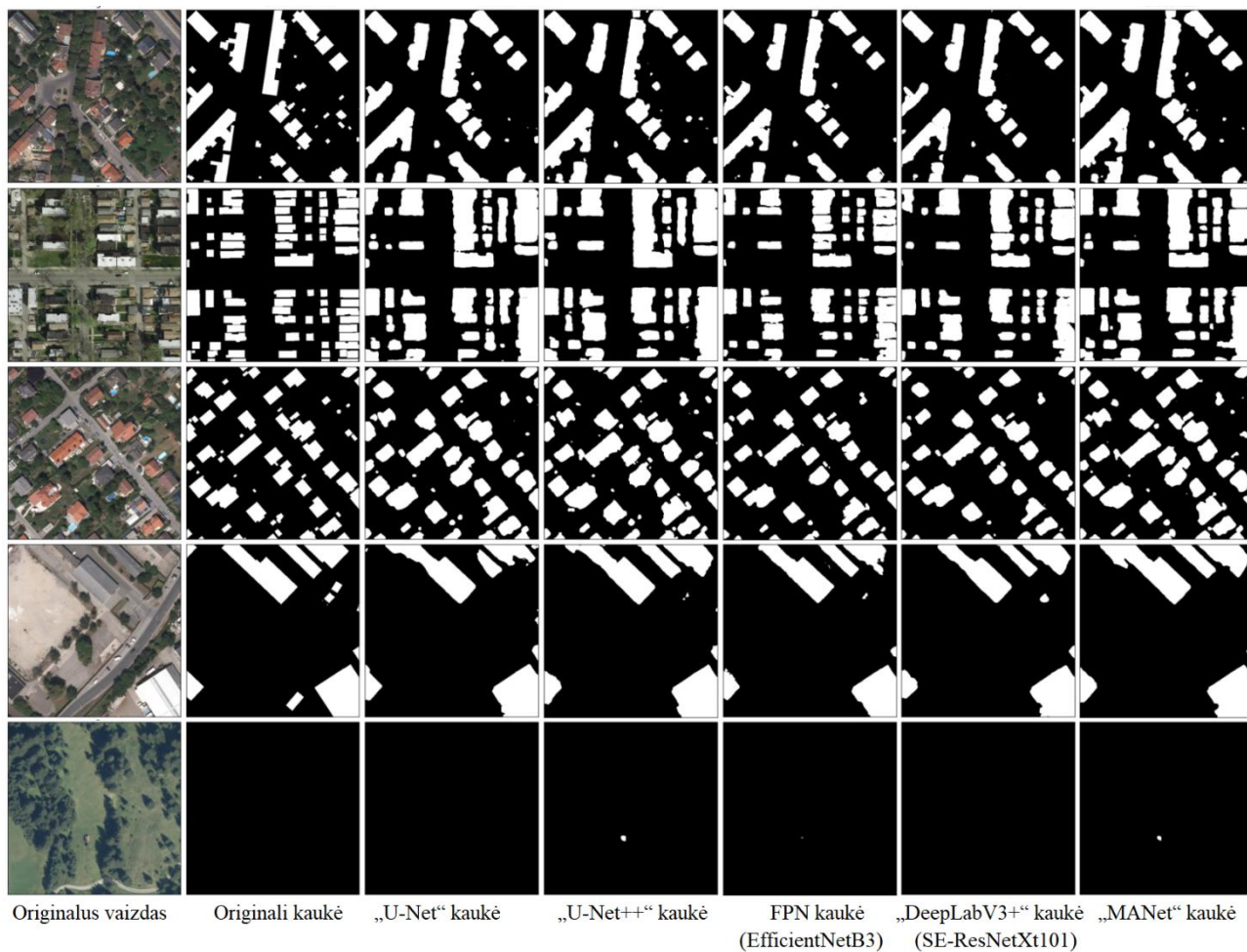


Originalus vaizdas

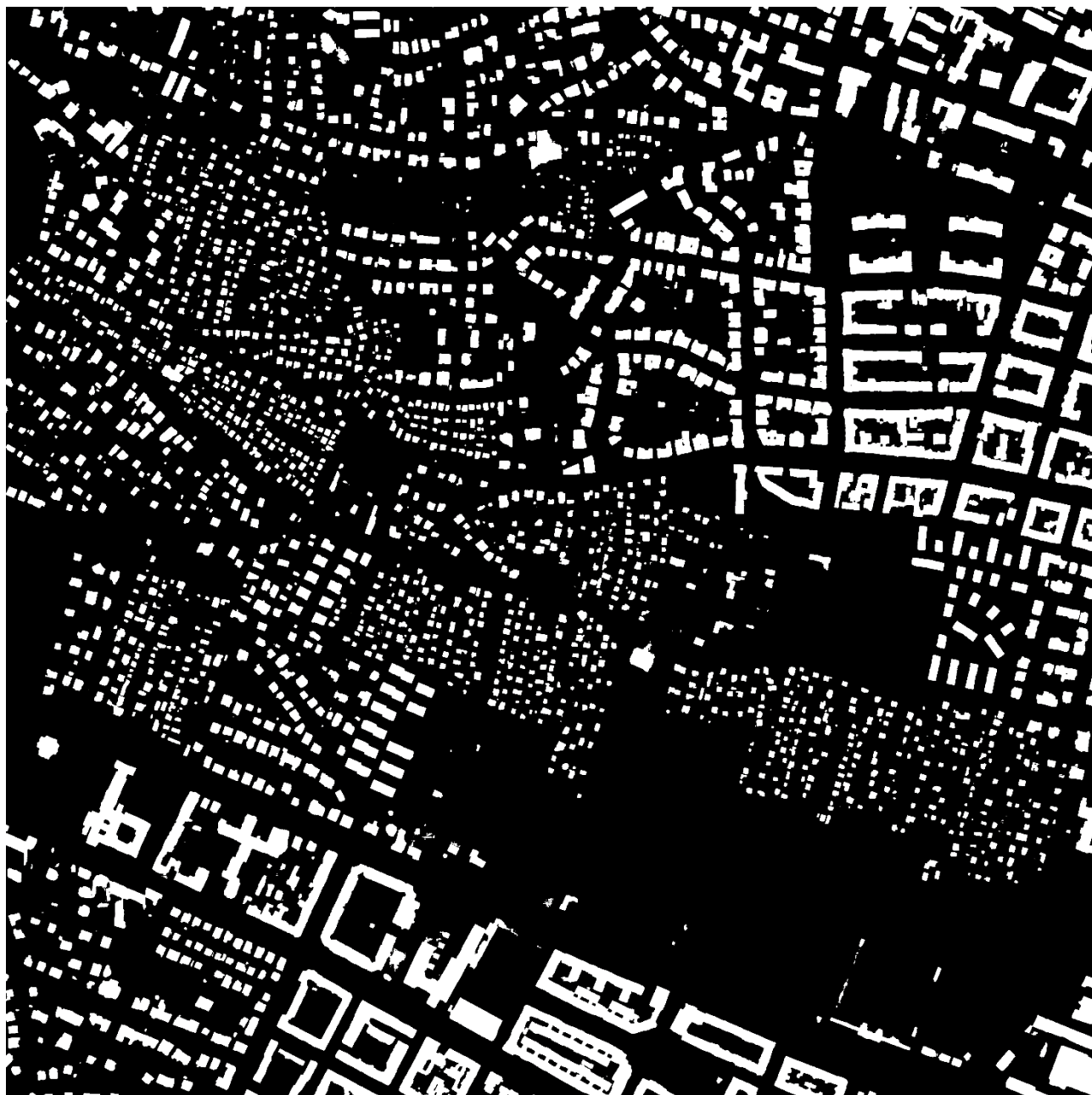
Originali kaukė

„DeepLabV3+“ kaukė

Šiame paveiksle pateikti Masačusetso pastatų duomenų rinkinio segmentavimo pavyzdžiai.



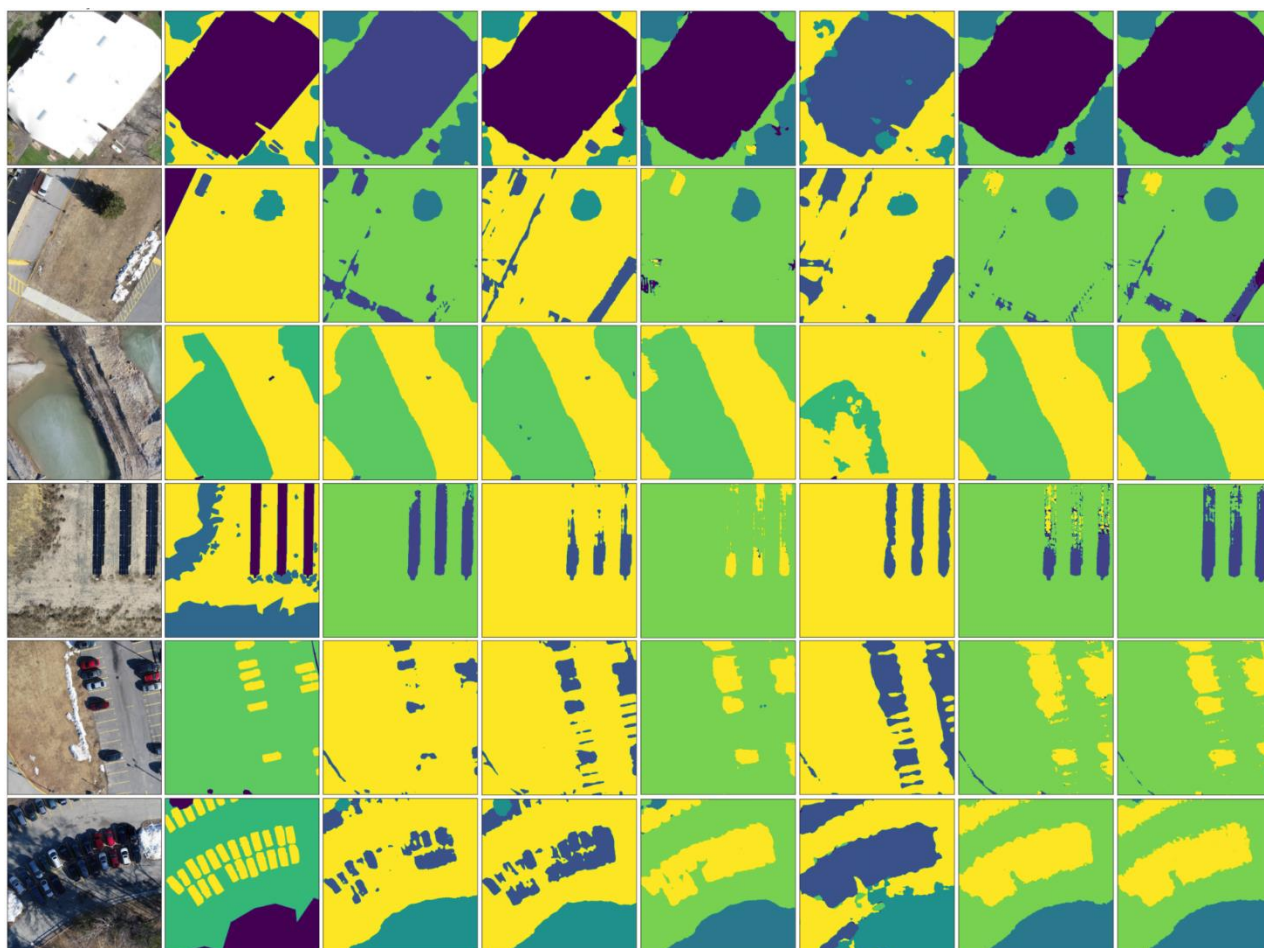
Šiame paveiksle pateikti „INRIA“ duomenų rinkinio segmentavimo pavyzdžiai.



Šiame paveiksle pateikta „INRIA“ rinkinio originalaus dydžio (5000×5000 pikselių raiškos) PPT kaukė, kuri buvo sujungta iš mažesnės raiškos dalių.



Šiame paveiksle pateikta „INRIA“ rinkinio originalaus vaizdo ir PPT modelio kaukės perdanga. Geltona spalva žymimi teisingai segmentuoti pikseliai, žalia spalva rodomi trūkstami tikrieji segmentai, o raudona spalva – klaidingai segmentuotos vietos.



Originalus vaizdas Originali kaukė „U-Net“ kaukė „U-Net++“ kaukė FPN kaukė „PSPNet“ kaukė „MANet“ kaukė „DeepLabV3+“ kaukė

Šiame paveiksle pateikiamos „DroneDeploy“ originalios nuotraukos, jų kaukės ir skirtingų modelių prognozuotos kaukės.

B priedas. Pranešimo XXIV-oje Lietuvos jaunųjų mokslininkų konferencijoje „MOKSLAS – LIETUVOS ATEITIS“ medžiaga

XXIV-toji Lietuvos jaunųjų mokslininkų konferencija
„MOKSLAS – LIETUVOS ATEITIS“
ELEKTRONIKA IR ELEKTROTECHNIKA

PAŽYMA
Vykintas Vyšniauskas

Vaizdų technologijų (T111) sekcijoje perskaitė pranešimą

***AUTOMATINĖ AEROFOTOGRAFINIŲ NUOTRAUKŲ
ANALIZĖ TAIKANT GILIOJO MOKYMOSI METODUS***

Mokslo komiteto pirmininkas

prof. habil. dr. Romanas Martavičius

Sekcijos pirmininkas

dr. Eldar Šabanovič



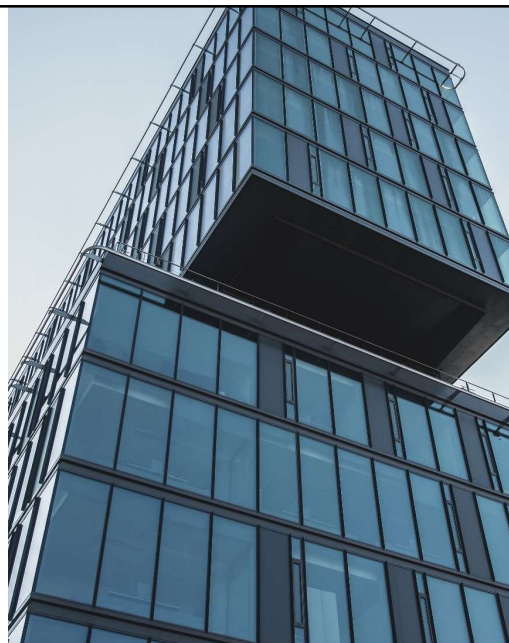
2021 m. kovo 19 d.
Vilnius

Vykintas Vyšniauskas DISfm-19

Automatinė aerofotografinių nuotraukų analizė taikant giliojo mokymosi metodus

Aktualumas ir problematika

Palydoviniai duomenys yra naudojami žvalgybai atlikti, miestams planuoti, augmenijai stebėti, gamtos ištekliams stebėti, klimato kaitos tyrimams atlikti, geologiniams pavojams stebėti, vandenynams stebėti ir daugelyje kitų sričių. Dėl šių vaizdų sudėtingumo susiduriama su tokiais iššūkiais kaip daugiamačiškumas, netiesiškumas, didelė tos pačios klasės objektų įvairovė, didelis skirtingų klasių objektų panašumas, didelis objekto ir scenos mastelių skirtumas ir kt. Dėl šių priežasčių dabartiniais analizės metodams vis dar trūksta tikslumo, todėl neišvengiamai reikia naudoti žmogaus intelektą priimančias sprendimus.



Automatinė aerofotografinių nuotraukų analizė taikant
giliojo mokymosi metodus



Dvejtainio semantinio segmentavimo tyrimai

1. Kelių išskyrimas

- Masačusetso kelių duomenų rinkinys
- Lietuvos miestų ir miestelių kelių duomenų rinkinys, sudarytas naudojant *beta.maps.lt* žemėlapius.

2. Pastatų išskyrimas

- Masačusetso pastatų duomenų rinkinys
- „INRIA Aerial Image Labeling“ duomenų rinkinys

Automatinė aerofotografinių nuotraukų analizė
taikant giliojo mokymosi metodus

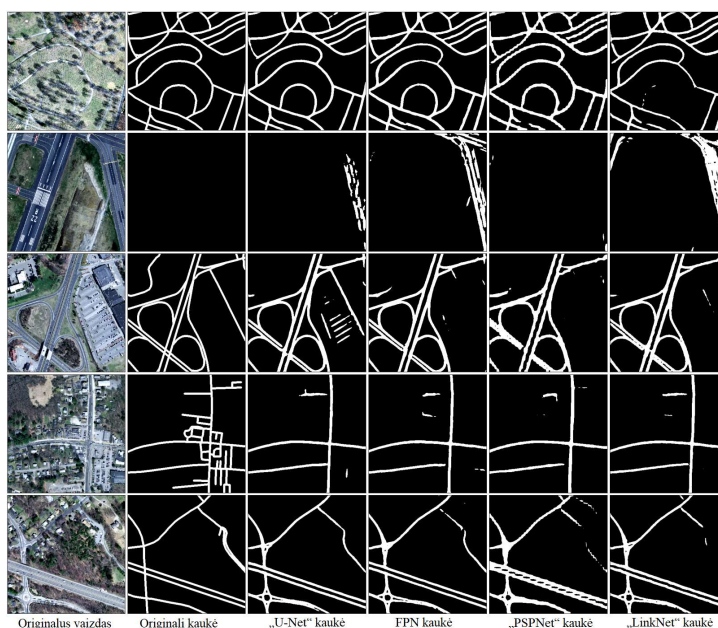
2021-03-19



Rezultatai

1 lentelė. Kelių segmentavimo rezultatai naudojant Masačusetso kelių duomenų rinkinį

Modelis	IoU, %	„F1“, %
„U-Net“	62,1	73,5
FPN	60,5	72,5
„PSPNet“	58,6	70,9
„LinkNet“	61	72,7



Originalus vaizdas

Originali kaukė

„U-Net“ kaukė

FPN kaukė

„PSPNet“ kaukė

„LinkNet“ kaukė

Automatinė aerofotografinių nuotraukų analizė
taikant giliojo mokymosi metodus

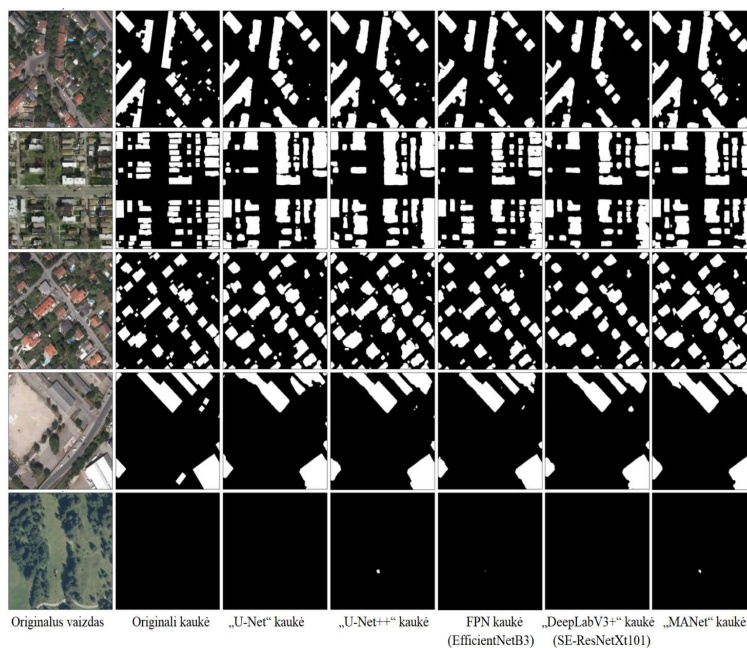
2021-03-19



Rezultatai

2 lentelė. Pastatų segmentavimo rezultatai naudojant „INRIA“ pastatų duomenų rinkinį

Modelis	IoU, %	„F1“, %
„U-Net“	67,82	77,43
„U-Net++“	66,99	76,51
FPN	75,47	83,33
„DeepLabV3+“	75,79	83,93
„MANet“	69,59	78,81



Originalus vaizdas Originali kaukė „U-Net“ kaukė „U-Net++“ kaukė FPN kaukė „DeepLabV3+“ kaukė „MANet“ kaukė
(EfficientNetB3) (SE-ResNetXt101)

Automatinė aerofotografinių nuotraukų analizė
taikant giliojo mokymosi metodus

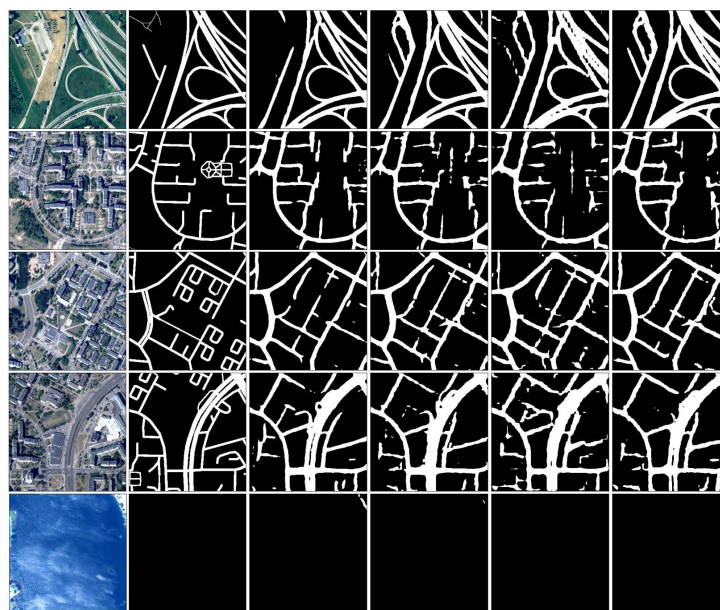
2021-03-19



Rezultatai

3 lentelė. Lietuvos kelių segmentavimo rezultatai

Modelis	IoU, %	„F1“, %
„U-Net“	41,8	57,3
FPN	42	57,5
„PSPNet“	41	56,5
„LinkNet“	41,6	57,2



Originalus vaizdas Originali kaukė „U-Net“ kaukė FPN kaukė „PSPNet“ kaukė „LinkNet“ kaukė

Automatinė aerofotografinių nuotraukų analizė
taikant giliojo mokymosi metodus

2021-03-19



Išvados

- Išankstinis apdorojimas (angl. Pre-processing)
- Apdorojimas klasikiniais metodais, pvz., CRF (angl. Conditional Random Fields)
- Tikslo funkcijų modifikacijos
- Duomenų papildymas



Ačiū už dėmesį!

