

Debesų šalinimas iš ortografinių nuotraukų naudojant giliuosius generatyvinius tinklus

Dominykas Kaminskas

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Didlaukio g. 47, LT-08303 Vilnius
dominykas.kaminskas@mif.stud.vu.lt

Santrauka. Debesys daro neigiamą poveikį Žemės stebėjimui ir palydovų užfiksuojamiems vaizdams. Šiame darbe analizuojami metodai, skirti debesų pašalinimui iš RGB (regimosios šviesos) tipo nuotraukų. Tyrimo metu 3 giliaisiais generatyviais tinklais paremti trukdžių šalinimo metodai – *CloudGAN* [1], *LaMa* [2] ir *RePaint* [3] – pritaikyti debesuotam Vilniaus miesto ortografinių (palydovinių) nuotraukų rinkiniui. Šių modelių sugeneruotos nuotraukos panaudotos trejetų neuroninio tinklo apmokymui. Šalinimo metodų efektyvumui nustatyti apskaičiuotos nuotraukų panašumo metrikos. Geriausiai pasirodė *LaMa* modelis, pasižymėjęs savo greitaveika, siūlomu funkcionalumu ir aukštais tyrimo rezultatais.

Raktiniai žodžiai: gilusis mokymas, generatyviniai tinklai, difuziniai modeliai, debesų šalinimas, palydovinės nuotraukos.

1 Įvadas

Debesys bet kuriuo laiko momentu dengia apie du trečdalius (67 %) Žemės paviršiaus [4]. Daugiau nei 50 % palydovų užfiksuotų vaizdų padengti debesų masyvais. Tai daro neigiamą poveikį Žemės stebėjimui iš kosmoso bei surinkamai duomenų kokybei. Šiai problemai spręsti atliekami įvairūs tyrimai, siekiantys sukurti debesų pašalinimo iš palydovinių nuotraukų metodus.

Tradiciniai debesų šalinimo algoritmai iš palydovinių nuotraukų grindžiami papildoma informacija – laiko eilučių [5] arba spektriniais [6] duomenimis. Tačiau papildomi duomenys gali būti nepasiekiami arba išvis neegzistuoti.

Gilieji neuroniniai tinklai per pastaruosius keletą dešimtmečių padarė nepaprastą pažangą [7]. Generatyvinio dirbtinio intelekto pagalba įmanoma kurti realistiškus vaizdus ir generuoti objektus ant nuotraukų paviršiaus.

Šiame darbe lyginami 3 giliaji generatyvinių tinklų metodai, pasižymėję savo efektyvumu šalinant trukdžius iš nuotraukų bei generuojant naujus vaizdus [8]. Atlikus literatūros analizę nuspręsta pritaikyti 2 bendro pobūdžio generatyvinius trukdžių šalinimo algoritmus – *LaMa* ir *RePaint* – bei išbandyti 1 debesų šalinimo užduočiai skirtą generatyvinį neuroninį tinklą – *CloudGAN*. Darbo tikslas – nustatyti efektyviausią metodą debesų šalinimo iš palydovinių nuotraukų užduočiai spręsti. Modelių efektyvumas palygintas naudojantis trejetų neuroniniu tinklu, gebančiu apskaičiuoti nuotraukų panašumo metrikas.

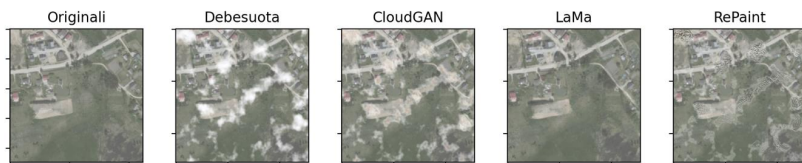
2 Gilieji generatyviniai tinklai

Generatyvinis dirbtinis intelektas yra vienas didžiausių 21 a. kompiuterių mokslo proveržių, pakeitęs žmonių profesinį, asmeninį ir socialinį gyvenimus [23]. Generatyviniai modeliai geba kurti tekstą, garsą ar net vaizdą, todėl gali būti pritaikyti duomenų analizei, nuotraukų generavimui ar kodo optimizacijai [9]. Šie modeliai taip pat gali būti panaudoti trukdžių šalinimui iš vaizdų (angl. *image inpainting*) - remiantis kontekstine nuotraukos informacija pageidaujamoje zonoje generuojamos naujos pikselių reikšmės.

Tyrimo metu išbandyti 3 generatyviniai modeliai:

- *CloudGAN* - generatyvinių besivaržančių tinklų (angl. *Generative Adversarial Network*) modelis, skirtas atpažinti ir pašalinti debesų masyvus iš palydovinių RGB tipo nuotraukų [1]. Išbandyta originali autorių implementacija, siūlanti patogias konsolines komandas programos vykdymui.
- *LaMa* - generatyvinių besivaržančių tinklų (GAN) modelis, leidžiantis pašalinti trukdžius bei įterpti pageidaujamus objektus nuotraukose [2]. Šiame darbe išbandytas *Python* aplinkos paketas "*simple-lama-inpainting*" [10], leidžiantis greitai instaliuoti programinę įrangą ir įvertinti jos efektyvumą.
- *RePaint* - difuzinis dirbtinio intelekto modelis [11], skirtas vaizdų atkūrimo nuotraukose užduotims atlikti [3]. Panaudota kompanijos *Hugging Face* algoritmo implementacija iš "*Diffusers*" bibliotekos [12].

Trukdžių šalinimo algoritmų rezultatą galima vizualiai palyginti 1 paveikslėlyje.



1 pav. Pateiktos originalios ir debesuotos nuotraukos bei skirtingų trukdžių šalinimo algoritmų (*CloudGAN*, *LaMa*, *RePaint*) išvestys.

Visi šiame darbe išbandyti generatyviniai modeliai surašyti 1 lentelėje. Pateiktas vykdymo laikas, skirtas vienai nuotraukai apdoroti. Galima pastebėti, jog *CloudGAN* ir *LaMa* modeliai savo našumu smarkiai lenkia *RePaint* algoritmą. Taip yra todėl, nes difuziniai modeliai paremti iteratyviu veikimo principu, susidedančiu iš nykstamojo proceso (angl. *forward process*) ir grįžtamojo proceso (angl. *backward process*) [11]. Difuzinė architektūra reikalauja daugiau operacijų skaičiaus nei generatyviniai besivaržančių tinklų (GAN) modeliai tai pačiai užduočiai įvykdyti.

1 lentelė. Modelių palyginimas.

Modelis	Tipas	Vykdymo laikas, sek	Debesų aptikimas	Nuotraukos dydis, px
<i>CloudGAN</i>	GAN	0,8	Taip	256 x 256
<i>LaMa</i>	GAN	0,1	Ne	Be apribojimų
<i>RePaint</i>	Difuzinis	11	Ne	256 x 256

CloudGAN yra vienintelis šiame darbe aptariamas modelis, gebantis automatiškai aptikti debesis – generuoti debesų kaukes (angl. *cloud masks*). Nenurodžius debesų sluoksnio, algoritmas bandys pats atpažinti debesų teritorijas ir jas pašalinti, o nuotraukos zonas, kurias modifikavo, pažymės sugeneruotame debesų sluoksnyje. Tačiau šis funkcionalumas tyrimo metu pasirodė neefektyvus – spalvos iškreipiamos, paveikslėlis tampa neatpažįstamas (sugadintas). Todėl galima teigti, jog efektyviai pašalinti debesis be trukdžių sluoksnio (angl. *inpaint mask*) su darbe aprašomais metodais neįmanoma.

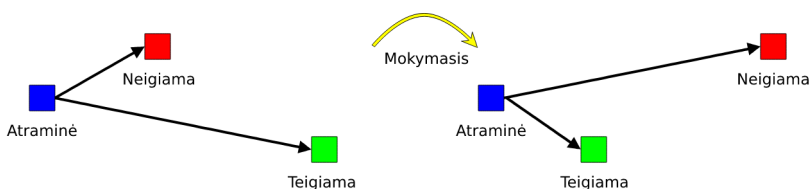
Vienintelis šiame darbe išbandytas modelis, įvesčiai priimantis įvairaus dydžio nuotraukas (netaikantis apribojimų), yra *LaMa*. Nors tinklas buvo apmokytas tik su 256 x 256 px rezoliucijos vaizdais, algoritmas sėkmingai

pritaikomas ir didesnės rezoliucijos nuotraukoms. Autorių teigimu generatyvinis tinklas sėkmingai veikia ir su nuotraukų rezoliucijomis, siekiančiomis 2000 x 2000 px. Likę modeliai (*CloudGAN* ir *RePaint*) įvesčiai priima tik standartinių matmenų vaizdus, todėl visame darbe nuspręsta taikyti vieną rezoliuciją (256 x 256 px).

3 Trejetų neuroninis tinklas

Nuotraukų panašumo metrikos naudojamos algoritmų efektyvumui įvertinti [13]. Šiame darbe pritaikytas gilusis konvoliucinis tinklas palydovinių vaizdų panašumui nustatyti (toliau – trejetų neuroninis tinklas). Sėkmingam nuotraukų panašumo įvertinimui pasitelkta trejetų nuostolių funkcija (angl. *triplet loss*) [14].

Trejetų neuroninio tinklo mokymui skirtą duomenų rinkinį visada sudarys nuotraukų trejetai: atraminė nuotrauka (angl. *anchor image*), teigiama nuotrauka (angl. *positive image*) ir neigiama nuotrauka (angl. *negative image*). Tikslas - mažinti atstumą tarp atraminės-teigiamos nuotraukų poros tuo pačiu metu didinant atskirtį tarp atraminės ir neigiamos nuotraukų (2 paveikslėlis). Šių atstumų skirtumas (toliau - vidutinis skirtumas) nurodo tinklo efektyvumą skirti panašias nuotraukas nuo nepanašių.



2 pav. Trejetų neuroninio tinklo mokymo metu siekiamas tikslas.

Ortografinių nuotraukų panašumo požymiams nustatyti buvo pasirinktos 5 neuroninių tinklų architektūros: *MobileNet* [15], *EfficientNet* [16], *EfficientNetV2* [17], *VGG-16* [18] ir *ResNet* [19], bei jų įvairios konfigūracijos. Vilniaus universiteto mokslininkai savo darbe nustatė, jog naudojant *VGG-16* architektūrą kartu su trejetų nuostolių funkcija galima pasiekti iki 99 % tikslumą [20], todėl šis modelis laikytinas standartu.

4 Darbo eiga

Toliau aprašoma detali tyrimo eiga:

1. Iš Vilniaus miesto ortografinių žemėlapių nuotraukų [21] buvo sudarytas duomenų rinkinys, susidedantis iš daugiau kaip 20 tūkst. nuotraukų. Dirbtinio intelekto modeliui apmokyti pasirinkta 60 % visų nuotraukų, o validacijai ir testavimui skirta po 20 %.
2. Duomenų rinkiniui buvo pritaikytas dirbtinių debesų generavimo įrankis *Satellite Cloud Generator* [22]. Tai parametrizuota *Python* programa, leidžianti ant nuotraukų paviršiaus generuoti realistiškai atrodančius debesis.
3. Debesuotoms nuotraukoms pritaikyti trukdžių šalinimo algoritmai (*CloudGAN, LaMa, RePaint*) debesims panaikinti.
4. 5 kartus atskirai apmokytas trejetų neuroninis tinklas trukdžių šalinimo metodų efektyvumui įvertinti pritaikius:
 - a. Originalų duomenų rinkinį;
 - b. Debesuotų nuotraukų rinkinį;
 - c. *CloudGAN* pašalintų debesų nuotraukų rinkinį;
 - d. *LaMa* pašalintų debesų nuotraukų rinkinį;
 - e. *RePaint* pašalintų debesų nuotraukų rinkinį.
5. Apskaičiuotos nuotraukų panašumo metrikos (vidutinis skirtumas ir tikslumas) bei palyginti rezultatai.

5 Rezultatai

Kiekvieną kartą tinklui baigus mokymosi procesą buvo vykdomi šie eksperimentai:

- Greitaveikos eksperimentas – skaičiuojamas laikas, per kurį neuroninis tinklas ortografines nuotraukas paverčia požymių žemėlapiais (angl. *feature maps*).
- Vidutinio skirtumo eksperimentas - trejetų neuroninio tinklo tikslo funkcijos reikšmė pateikus testavimui skirtą nuotraukų rinkinį (mokymo metu nematyti vaizdai).
- Tikslumo eksperimentas – tinklo gebėjimas pateikus nuotrauką rasti panašias iš didelės krūvos vietovės vaizdų.

Generatyvinio modelio efektyvumas šalinant debesis iš ortografinių nuotraukų vertinamas pagal tai, kiek tinklas eksperimentų metu prano-

ko debesuotų nuotraukų rinkinio pasiektus rezultatus ir kiek atsiliko nuo originalių nuotraukų (angl. *ground truth*). Trejetų neuroninio tinklo VGG-16 architektūros eksperimentų rezultatai pateikti 2 lentelėje. Paryškintos tos trukdžių šalinimo algoritmų reikšmės, kurios pranoko debesuotų nuotraukų rinkinio pademonstruotus rezultatus.

2 lentelė. Trejetų neuroninio tinklo VGG-16 architektūros pasiekti rezultatai.

Nuotraukų rinkinys	Laikas, sek ↓	Vidutinis skirtumas ↑	Tikslumas, % ↑
Originali	13	0,22	99,18
Debesuota	21	0,17	88,62
<i>CloudGAN</i>	17	0,15	87,16
<i>LaMa</i>	13	0,21	90,46
<i>RePaint</i>	13	0,15	73,22

Debesų daroma įtaka nuotraukų panašumo nustatymui – vidutiniškai apie 5-10 % kritęs tikslumas. Sėkmingai pasirodęs *LaMa* modelis pranoko debesuotų nuotraukų rinkinį visose panašumo metrikose ir pasiekė beveik 99 % tikslumą su *ResNet* architektūra, todėl yra efektyviausias išbandytas algoritmas šiame darbe šalinant debesis iš ortografinių nuotraukų. Šio modelio sugeneruotos nuotraukos beveik visais atvejais pademonstravo apie 2-5 % didesnį tikslumą už debesuotus vaizdus. Ne toks efektyvus buvo *CloudGAN* modelis, tyrimų metu pranokęs debesuotų duomenų rinkinį tik 7 iš 18 kartų (apie 40 %). Prasčiausiai pasirodė *RePaint* modelis, kurio pašalintų nuotraukų rinkinys eksperimentų metu vos kelis kartus pagerino debesuotų nuotraukų panašumo metrikas. Žinant, jog bandymų metu šio modelio vykdymo laikas buvo lėčiausias, būtų galima teigti, jog *RePaint* nėra efektyvus algoritmas šalinant debesis iš ortografinių nuotraukų.

6 Išvados

Debesų poveikis palydoviniam Žemės stebėjimui ar nuotoliniams moksliniams tyrimams – prarasta dalis surenkamos informacijos. Tradiciniai metodai duomenų atkūrimui pagrįsti papildomais informacijos šaltiniais - laiko eilučių ar spektriniais. Tačiau papildomi duomenys ne visada egzistuoja. Todėl šiame darbe palyginti generatyvinių tinklų modeliai, gebantys pašalinti dalį debesų iš ortografinių nuotraukų. Tyrimų metu geriausiai pasirodė *LaMa* algoritmas, kurio sugeneruotų nuotraukų rinkinio rezultatai panašu-

mo metrikomis (vidutiniu skirtumu ir tikslumu) aplenkė debesuotų nuotraukų rinkinio rezultatus. Prasčiausiai pasirodė *RePaint* algoritmas, nepasižymėjęs nei vykdymo laiku, nei panašumo metrikomis. Taip pat šis tyrimas parodė, jog nėra efektyvaus būdo pašalinti debesų iš RGB tipo palydovinių nuotraukų neturint debesų kaukės (angl. *cloud mask*).

7 Padėka

Padėka skirta Vilniaus universiteto Matematikos ir informatikos fakulteto Informacinių technologijų atviros prieigos centrui (ITAPC) už suteiktus HPC išteklius šio darbo skaičiavimams atlikti. Eksperimentai buvo vykdomi paskirstytų skaičiavimų tinkle (PST) - tai 2021 metais pristatytas naujasis universiteto superkompiuteris, kurio bendras teorinis našumas yra ne mažesnis nei 0,5 Pflops DP.

Literatūra

- [1] Kluiver, S., Schonenberg, J. (2022). CloudGAN: Detecting and Removing Clouds from RGB-images using Image Inpainting.
- [2] Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., ... & Lempitsky, V. (2022). Resolution-robust large mask inpainting with fourier convolutions. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 2149-2159).
- [3] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11461-11471).
- [4] King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., & Hubanks, P. A. (2013). Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. *IEEE transactions on geoscience and remote sensing*, 51(7), 3826-3852.
- [5] Alonso-Sarria, F., Valdivieso-Ros, C., & Gomariz-Castillo, F. (2024). Imagery Time Series Cloud Removal and Classification Using Long Short Term Memory Neural Networks. *Remote Sensing*, 16(12), 2150. <https://doi.org/10.3390/rs16122150>
- [6] Hasan, C., Horne, R., Mauw, S., & Mizera, A. (2022). Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks. *International Journal of Remote Sensing*, 43(5), 1881–1893. <https://doi.org/10.1080/01431161.2022.2048915>
- [7] Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial intelligence in the 21st century. *Ieee Access*, 6, 34403-34421.
- [8] Liu, Z. (2024). Inpainting Ortho-Rectified Satellite Images. *Photogrammetry and Remote Sensing Institute of Geodesy and Photogrammetry, ETH Zürich*.
- [9] Gupta, R., Nair, K., Mishra, M., Ibrahim, B., & Bhardwaj, S. (2024). Adoption and impacts of generative artificial intelligence: Theoretical underpinnings and research agenda. *International Journal of Information Management Data Insights*, 4(1), 100232.

- [10] Simple LaMa Inpainting. (2022). <https://github.com/enesmsahin/simple-lama-inpainting>
- [11] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33, 6840-6851.
- [12] Platen, P., Patil, S., Lozhkov, A., Cuenca, P., Lambert, N., Rasul, K., Davaadorj, M., Nair, D., Paul, S., Berman, W., Xu, Y., Liu, S., & Wolf, T. (2022). Diffusers: State-of-the-art diffusion models. <https://github.com/huggingface/diffusers>
- [13] Mitchell, H.B. (2010). Image Similarity Measures. In: *Image Fusion*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-11216-4_14
- [14] F. Schroff, D. Kalenichenko, J. Philbin. Facenet: A unified embedding for face recognition and clustering. (2015). *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 815–823).
- [15] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [16] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [17] Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.
- [18] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- [19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [20] Valaitis, V., Marcinkevičius, V., & Jurevičius, R. (2020). Learning aerial image similarity using triplet networks. In *Numerical computations: theory and algorithms: international conference on numerical computations: theory and algorithms NUMTA 2019. Part 2 / edited by: Sergeyev, YD; Kvasov, DE*. https://doi.org/10.1007/978-3-030-40616-5_15
- [21] V. Ž. ūkio duomenų centras. (2025). Lietuvos erdvinės informacijos portalas. <https://www.geoportal.lt/geoportal/>
- [22] Czerkawski, M., Atkinson, R., Michie, C., & Tachtatzis, C. (2023). SatelliteCloudGenerator: Controllable Cloud and Shadow Synthesis for Multi-Spectral Optical Satellite Images. *Remote Sensing*, 15(17), 4138. <https://doi.org/10.3390/rs15174138>
- [23] Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 63.