

# Laukų segmentavimas iš palydovinių nuotraukų

Aurimas Arlauskas

*Informatikos institutas*

*Matematikos ir informatikos fakultetas*

Vilnius, Lietuva

aurimas.arlauskas@mif.stud.vu.lt

Tautvydas Mačiulis

*Informatikos institutas*

*Matematikos ir informatikos fakultetas*

Vilnius, Lietuva

tautvydas.maciulis@mif.stud.vu.lt

Andrius Bušma

*Informatikos institutas*

*Matematikos ir informatikos fakultetas*

Vilnius, Lietuva

andrius.busma@mif.stud.vu.lt

Julius Rasimas

*Informatikos institutas*

*Matematikos ir informatikos fakultetas*

Vilnius, Lietuva

julius.rasimas@mif.stud.vu.lt

**Santrauka**—Šiame darbe atliekamas modelių tikslumo palyginimas naudojant palydovines ūkio paskirties laukų nuotraukas. Tyrimo metu naudojamas “Agriculture-Vision” duomenų rinkinys, kuriuo yra apmokomi modeliai “FPN”, “UNet” ir “DeepLabV3” objektų segmentacijos užduočiai. Geriausią mIoU metriką turėjo modelis FPN.

**Index Terms**—feature, FCN, UNet, FPN, DeepLabV3, agriculture, mIoU

## I. ĮVADAS

Mašininė rega yra dirbtinio intelekto šaka, kuri leidžia programoms sudaryti reikšmingą informaciją iš nuotraukų, paveikslėlių, video medžiagos, ar kitų vizualinių šaltinių. Vienas iš mašinės regos uždavinių yra paveikslėlių segmentavimas.

Paveikslėlių segmentavimas yra procesas, kuriame paveikslėlis yra padalinamas į keletą segmentų (pikselių aibių). Šio proceso metu kiekvienam pikseliui yra priskiriama etiketė (angl. *label*) pagal kažkokiais charakteristikas. Paveikslėlių segmentavimo užduoties rezultatas yra segmentų aibė, kurių pikseliai užkloja visą paveikslėlį pilnai (t.y. kiekvienas pikselis yra priskirtas kažkuriam segmentui (objektui)). Galimų objektų aibė dažnu atveju susidaro iš norimų rasti objektų bei fonų. [1]

Mašinės regos pritaikymas agrikultūroje pradėjo plėstis išsiplėtus giliojo mokymosi metodų pajėgumams spręsti įvairias užduotis. To pasekoje daug pastangų pradėta dėti į segmentavimo pritaikymą agrikultūroje, norint padidinti auginamų kultūrų derlių bei sumažinti ūkio kaštus. Vis dėlto, substantyvūs rezultatai šioje srityje dar nėra pasiekti [5], kas gali būti paaiškinama duomenų rinkinių (angl. *dataset*) trūkumu.

Semantinė segmentacija iš palydovinių nuotraukų yra viena pagrindinių mašinės regos pritaikymo agrikultūroje būdų. Ūkinės paskirties laukuose vyrauja įvairiausios objektų formos ir dydžiai. Dėl šio ir kitų aspektų palydovinių agrikultūrinių nuotraukų segmentacija kelia ypač daug sunkumų, bet taip pat turi didžiulį akademinį bei ekonominį potencialą [3].

## II. TIRIAMY MODELIAI

Šiame darbe tyrimui buvo pasirinkta naudoti tris modelių architektūras: UNet, FPN ir DeepLabV3. Kiekvienai architektūrai koduotojo (angl. *encoder*) sluoksniui naudotas ResNet18 konvoliucinis neuroninis tinklas, kuris buvo ištreniruotas su ImageNet duomenų rinkiniu. Visos naudotos architektūros buvo paimtos iš viešai prieinamos Github repozitorijos.

### A. UNet

U-Net (UNet) yra pilnai konvoliucinis neuroninis tinklas (toliau *FCN*), kuris buvo sukurtas biomedicininį nuotraukų segmentacijai. UNet sugeba tiksliai lokalizuoti transponuotas konvoliucijas. Tinklas susideda iš sutraukiančio ir išplečiančio takų — tai duoda modeliui U raidę primenančią architektūrą.

Sutraukiantis takas yra tipinis konvoliucinis tinklas, kuriame daug kartų naudojama konvoliucijos operacija, o po jos ReLU bei *max pooling* operacijos.

Išplečiantis takas sudeda *feature* ir erdvinę informaciją naudodamas konvoliucijas ir konkatenacijas su aukštos rezoliucijos *features* iš sutraukiančio tako. [9]

### B. FPN

FPN (Feature Pyramid Network) — FCN, kuriame sudedamos žemos rezoliucijos, semantiškai stiprios *features* ir aukštos rezoliucijos, semantiškai silpnos *features*. Modelis susideda iš apačios-į-viršų ir iš viršaus-į-apačia takų. Dėl vis mažėjančių erdvinų *features* modelio architektūra primena piramidę. [7]

### C. DeepLabV3

DeepLabV3 yra FCN. Vienas iš iššūkių naudojant FCN su paveikslėliais segmentacijos užduočiai yra, kad įeities *feature maps* tampa mažesni, judant per konvoliucinius ir *pooling* tinklo sluoksnius. Dėl to informacija apie paveikslėlius yra prarandama ir rezultatų spėjimai (angl. *predictions*) tampa žemos rezoliucijos bei objektų ribos nėra aiškios.

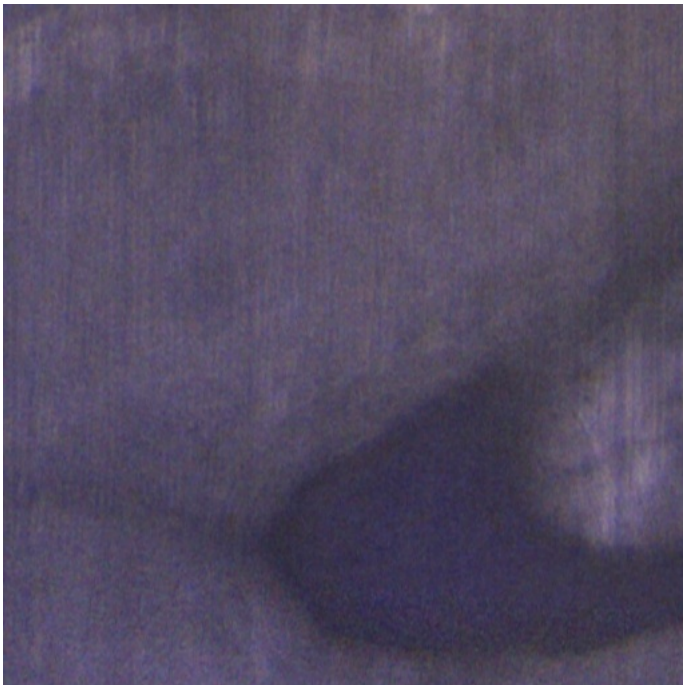
DeepLab modelis yra skirtas susitvarkyti su šiuo iššūkiu naudojant "Atrous" konvoliucijas bei "Atrous Spatial Pyramid Pooling (ASPP)" modulius.

DeepLabV3 išsiskiria nuo praeitų DeepLab versijų tuo, kad jo ASSP modulis yra pagerintas jame pridedant partijos normalizaciją (angl. *batch normalization*) bei paveikslėlio lygio *features* [2].

### III. DUOMENYS

Modeliams treniruoti ir testuoti pasirinkta naudoti Agriculture-Vision duomenų rinkinio [4] porinkinį. Porinkinį sudaro trys klasės: piktžolių klasteriai, vanduo ir maistinių medžiagų trūkumas. Kiekvieną klasę sudaro palydovinės žemės laukų  $512 \times 512$  pikselių dydžio nuotraukos, kuriose gali būti kiekvienos klasės laukų segmentai, t.y. jei tarkim nuotrauka yra iš vandens klasės, tai joje gali būti žemės lauko segmentas apsemtas vandens [1 pav]. Taip pat kiekviena nuotrauka turi atitinkamą etiketės (angl. *label*) nuotrauką, kurioje yra pavaizduota tikrosios tiesos (angl. *ground truth*) kaukė [2 pav]. Ši kaukė nurodo kurie konkretūs pikseliai priklauso atitinkamai klasei.

Agriculture-Vision duomenų rinkinyje piktžolių klasterių, vandens ir maistinių medžiagų trūkumo klasių treniravimui skirtų elementų yra atitinkamai po 11 tūkst., 2 tūkst. Ir 13 tūkst. Tam, kad išvengtų klasių disbalanso problemos, šiam tyrimui naudotame porinkinyje buvo pasirinkta sumažinti piktžolių klasterių ir maistinių medžiagų trūkumo klasių treniravimo elementų kiekį iki to paties kiekio, kiek yra vandens klasėje. Testavimo duomenų rinkinį per visas klases sudarė 10647 elementų.



1 pav. – Vandens klasės nuotraukos pavyzdys.



2 pav. – Vandens klasės nuotraukos kaukės pavyzdys.

### IV. TRENIRAVIMAS

Treniravimo metu buvo pasirinkta naudoti Dice loss nuostolių funkciją [8]. Tikslo funkcijos formulėje (1) nariai  $p_i$  ir  $g_i$  atitinkamai reiškia modelio numatytas (angl. *predicted*) ir tikrosios tiesos (angl. *ground truth*) pikselio reikšmes. Ši nuostolių funkcija buvo pasirinkta dėl to, kad ji skaičiavimo metu nuostolių informaciją apima ne tik lokaliai, bet ir globaliai, kas lemia didesnę tikslumą.

$$\mathcal{D} = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (1)$$

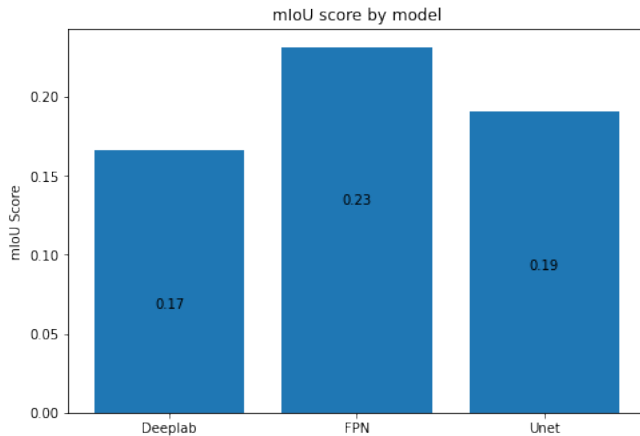
Kiekvienas modelis buvo treniruojamas 20 epochų, partijos (angl. *batch*) dydis buvo 2. Mokymosi greičio (angl. *learning rate*) reikšmė buvo pasirinkta 0.0001. Treniravimo metu buvo naudotas Adam optimizer [6] optimizacijos algoritmas. Modeliai buvo treniruoti naudojant Nvidia 980TI vaizdo plokštę.

### V. REZULTATAI

Ištreniravus šiame darbe tiriamus modelius, jie buvo paleisti su testavimo duomenimis. Tam, kad kiekvienas modelis būtų korektiškai palyginamas, buvo pasirinkta naudoti tą pačią metriką, kuri yra naudojama duomenų rinkinio konkurso (angl. *challenge*) metu [3]. Nurodytame konkurse yra naudojama mIoU metrika, kuri yra tiesiog suvidurkinta IoU reikšmė per visas tiriamas klases. Šios metrikos formulėje (2) nariai  $P_c$  ir  $T_c$  yra atitinkamai modelio numatytos ir tikrosios tiesos kaukių reikšmės.

$$mIoU = \frac{1}{c} \sum \frac{Area(P_c \cap T_c)}{Area(P_c \cup T_c)} \quad (2)$$

(3 pav.) pavaizduotoje diagramoje galime matyti modelių rezultatus. Geriausiai iš trijų modelių pasirodė FPN su 0.23 įverčiu. Antras pagal gerumą yra UNet su 0.19 mIoU ir galiausiai, trečioje vietoje, DeepLabV3 su 0.17.



3 pav. – Modelių mIoU rezultatai.

## VI. IŠVADOS

Šio tyrimo metu buvo atrasta, kad bendrinio tipo nuotraukų segmentavimo modeliai nesuteikia pakankamai gerų rezultatų žemės laukų segmentacijos uždavinyje. Tad galima daryti išvadą, kad yra būtinybė koreguoti esamus modelius, kad jie būtų labiau pritaikyti šiai užduočiai.

## LITERATŪRA

- [1] <https://www.ibm.com/topics/computer-vision>.
- [2] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.
- [3] Mang Tik Chiu, Xingqian Xu, Kai Wang, Jennifer Hobbs, Naira Hovakimyan, Thomas S. Huang, Honghui Shi, Yunchao Wei, Zilong Huang, Alexander Schwing, Robert Brunner, Ivan Dozier, Wyatt Dozier, Karen Ghandilyan, David Wilson, Hyunseong Park, Junhee Kim, Sungho Kim, Qinghui Liu, Michael C. Kampffmeyer, Robert Jenssen, Arnt B. Salberg, Alexandre Barbosa, Rodrigo Trevisan, Bingchen Zhao, Shaozuo Yu, Siwei Yang, Yin Wang, Hao Sheng, Xiao Chen, Jingyi Su, Ram Rajagopal, Andrew Ng, Van Thong Huynh, Soo-Hyung Kim, In-Seop Na, Ujjwal Baid, Shubham Innani, Prasad Dutande, Bhakti Baheti, Sanjay Talbar, and Jianyu Tang. The 1st agriculture-vision challenge: Methods and results, 2020.
- [4] Mang Tik Chiu, Xingqian Xu, Yunchao Wei, Zilong Huang, Alexander Schwing, Robert Brunner, Hrant Khachatrian, Hovnatan Karapetyan, Ivan Dozier, Greg Rose, et al. Agriculture-vision: A large aerial image database for agricultural pattern analysis. *arXiv preprint arXiv:2001.01306*, 2020.
- [5] Andreas Kamilaris and Francesc X Prenafeta-Boldu. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018.
- [6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [7] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection, 2017.
- [8] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016.
- [9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.