

Hatékony szőlő levelek szegmentálása drón felvételeken

Efficient segmentation of vine leaves on drone images

MOLNÁR Szilárd¹, dr. KERESZTES Barna², dr. TAMÁS Levente³

1-Kolozsvári Műszaki Egyetem, Szilard.Molnar@aut.utcluj.ro

2-Bordeaux-i Egyetem, Barna.Keresztes@ims-bordeaux.fr

3- Kolozsvári Műszaki Egyetem, Levente.Tamas@aut.utcluj.ro

Abstract

With the widening of the Agriculture 4.0 era, the use of autonomous robots in the agriculture field is becoming a priority. This is the case also in the vineyards where segmenting the vine canopies is a part of the pre-processing chain. This work presents a Feature Pyramid Network-based grape canopy segmentation method. We conducted our tests in different vineyards and obtained state-of-the-art segmentation results on public and custom datasets.

Keywords: applied robotics, agriculture, computer vision, artificial intelligence, precision viticulture

Kivonat

Az egyre bővülő számítógép által segített agrikultúrában (Agriculture 4.0), egyre nő az automatizált robotok szerepe. Ez a helyzet a szőlősökben is, ahol egy levélzet szegmentáló módszer nagy segítség lehet a későbbi folyamatokhoz. Ebben a munkában bemutatunk egy FPN-alapú levélzet szegmentáló módszert. További teszteket végzünk különböző adatszetekkel, illetve más szegmentáló algoritmusokkal szemben.

Kulcsszavak: alkalmazott robotok, mezőgazdaság, képfelismerés, mesterséges intelligencia, precíziós szőlőtermelés

1. BEVEZETÉS

Az IoT egyre nagyobb elterjedése több termelési ágazatban is paradigma váltást eredményez, így van ezzel az agrikultúra is, ahol egyre inkább megjelennek az automatizált robotok. Ezeknek a robotoknak az optimalizációja nagymértékű fejlesztést igényel, ezért több kutatócsoport próbál specializálódni egy adott terményre, mint például a szőlőre. Ha különböző számítástechnikai megoldásokat alkalmaznak a szőlősben, akkor azt precíziós szőlőtermelésnek vagy szőlőművelésnek nevezik.

Ez a terület önmagában is nagy befolyással bír, mivel a világon évente akár 30 millió tonna szőlőből is készíthetnek bort [3]. Ez nagy piaci részesedést eredményez, aminek köszönhetően az elmúlt harminc évben megszámú tudományos munka jelent meg precíziós szőlőtermelés témában ([4,10–12]).

A szőlősben több típusú robotizált járművet is felhasználhatnak. Az egyik kategória a földön járó robotok, amik nagy előnye a közeli, pontosabb vizsgálat. A másik nagyobb járműkategóriát a légi járművek képezik, mint például a drónok, amelyek a növényzet egységét tanulmányozva vizsgálják a lehetséges problémák jelenlétét. Ezekon az általánosításokon kívül természetesen drónnal is végezhetünk közeli megfigyelést, habár ez fokozott körülményt igényel. Sőt, egy drón használata, akár költséghatékonyabb is lehet, mint egy földi roboté, amennyiben egy drónra nincs hatással a talaj minősége, és általában nagyobb mozgási szabadsággal rendelkezik.

A közeli megfigyelés nagyobb részletességet biztosít, úgy az érdekelt növény, mint a környező növények esetében, ami akár meg is nehezítheti egy adott számítógépes látáson alapuló módszer dolgát. Emiatt, a kutatók igyekeznek leegyszerűsíteni a problémát azzal, hogy az érdekelt növényzetet meghagyják, míg minden más zavaró tényezőt kivágnak a képből. Ezt a folyamatot nevezik szegmentációnak. Ilyen szegmentáció nagy segítséggel lehet a navigációban [2], a termés megbecslésében [5] vagy a betegségek felismerésében [6].

Szegmentációt ígéretes lehet mélységi tanulási módszerekre alapozni. A szőlősben a levélzet több különböző méretű részt kell figyelembe venni, mint például kis és nagy levelek, indák, ágak vagy szőlőfürtök. Az architektúra kiválasztásánál figyelembe kell választani az előforduló változatosságot. Ilyen változatosságot

megfigyelhetünk, akkor is ha a térben számolunk felületi merőlegeseket, mivel ezek gyakran igényelnek különböző tartó felületeket. Ez a témája a [9] munkának, és ez adta az ihletet, hogy a használt architektúra egy Jellemző Piramis Háló (Feature Pyramid Network - FPN [8]) legyen.

Így ebben a dolgozatban bemutatunk egy FPN-alapú szegmentációs módszert, ami közeli mérésekből származó képeken szegmentálja ki a szőlő leveleit, amelyek megkönnyítik a további adatfeldolgozást. Továbbá szót ejtünk az adatokról, amelyeket keverve szereztünk földi és légi járművekről. Végül más módszerekkel is összehasonlítjuk a leírt módszert.

2. FELHASZNÁLT ADATOK ÉS ALGORITMUSOK

2.1. Adatszett

A szegmentációhoz olyan közeli képekre van szükségünk, amelyekhez tartoznak bináris maszkok, amelyek leírják, hogy a képen melyik képpont tartozik a levélzethez, és melyik nem.

Az első adatszettet, amit figyelembe vettünk [2] szolgáltatatta. Ez az észak-olaszországi adatszett 500 képet tartalmaz, mindegyikhez a párosítva a megfelelő bináris maszk.

Ehhez adtuk hozzá a saját képeinket, amelyet egy DJI Mini 2-es drónnal készítettünk. A helyszíneket a Kolozsvári Agronómia Egyetem szőlősei alkották Kolozsváron és a Szeben megyei Nagyapoldon. Ezekhez a képekhez természetesen manuálisan kellett elkészíteni a bináris képeket, ezért ebben a dolgozatban csupán 100 darab kép származik a mi adatszettünkben.

A két adatszettet összekevertük, majd választhatunk, hogy átméretezzük őket 640x480-as méretre, vagy ekkora méretűre daraboljuk fel az eredeti nagy felbontású képeket. Végül 1 az 5-höz arányban elosztottuk őket a teszt és tréning feladatokhoz. Példák a felhasznált képekre az 1. ábrán láthatóak.



1. ábra. Példa az adat szettre. Aghi (bal), saját (jobb).

2.2. Jellemző Piramis Háló (Feature Pyramid Network - FPN)

Ahogy korábban is említettük, ennek a dolgozatnak az alapja [9]. Elsőre megtévesztő lehet az a tény, hogy egy 2 dimenziós problémához a megoldást egy 3 dimenziós adatfeldolgozási munkából vesszük kölcsön, másodikkra viszont észrevehetjük, hogy a felhasznált FPN [8] architektúra kifejezetten alkalmas különböző méretű elemek tanulmányozására. És noha az említett dolgozatban a felületi normálisok számításához elsősorban mélységképek voltak használatban, az RGB-s bemeneti képek nem okoztak problémát.

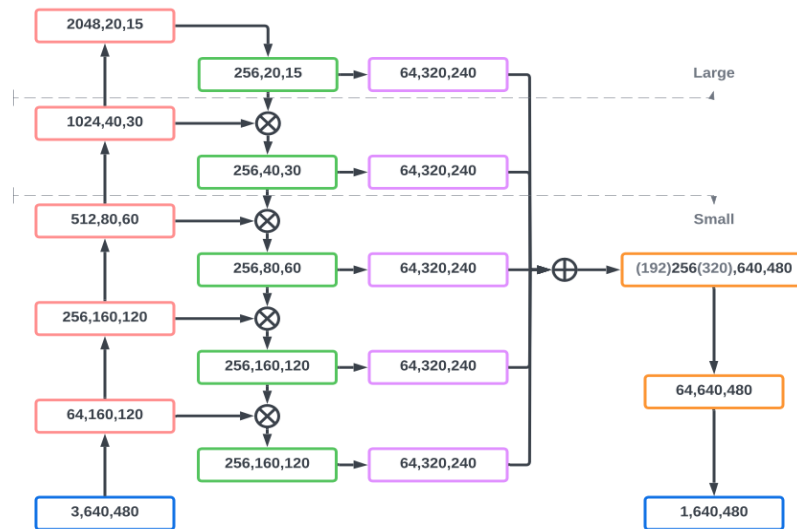
Az FPN architektúrát tulajdonképpen úgy is tekinthetjük, mint egy UNet modellt, ahol a két ág közötti kapcsolatokra további konvolúciókat illesztünk. Az így nyert architektúrában a baloldali ágot a ResNet101-re alapozzuk, ezáltal megkönnyítve a tanítást. A Pytorch alapú architektúra a 2. ábrán látható. Bemenetként megkap egy RGB képet, majd a háló végén a szigma réteg visszaad egy maszkot, amiből küszöböléssel egy bináris maszkot kapunk.

A tanításnál a modell célja, hogy maximalizálja a valós maszk és a kiszámolt maszkok közötti átfedést, azaz az Intersection over Union (IoU) metrikát (1):

$$IoU = \frac{X_{GT} \cdot X_{Pred}}{\sum_{i,j} (X_{GT}^{i,j} + X_{Pred}^{i,j} - X_{GT}^{i,j} * X_{Pred}^{i,j})} \quad (1)$$

ahol X_{GT} a valós maszk, míg X_{Pred} a kiszámolt maszk.

A tanítást egy Nvidia A100-as videokártyán végeztük, ezért nem okozott problémát a nagy modell méret, viszont azért, hogy más eszközökön is alkalmazható legyen ez a módszer, az architektúra mérete tetszés szerint állítható 3 különböző méret között. Ezek a méretek 6, 36 és 62 millió paramétert jelentenek, amikből



24, 140 illetve 250 MB-os súlyokat kaphatunk.

2. ábra. Az FPN architektúra. A szaggatott vonal jelöli a különböző méretek közötti tagolást.

3. KIÉRTÉKELÉS

A módszerünket mindhárom lehetséges mérettel teszteljük, illetve további két másik módszert is megemlítettük az összehasonlításnál, amelyeket gyakran használnak szegmentálási célokra. Ezek a Mask R-CNN [7] és a MobileNetV3, amelyet a [2] dolgozattól vettünk kölcsön. A saját modellek méretei korábban említettük, ezek mellett a Mask R-CNN mérete 290MB, a MobileNetV3-é pedig csupán 4MB, ami ígéretesnek tűnik beágyazott rendszerekhez. Mindegyik architektúrát ugyanazonokon a képeken tanítjuk be, illetve értékeljük ki. A kiértékelésnél az elsődleges szempontok a helyesen meghatározott képpontok aránya. Ezek mellett számolunk a hamis pozitív (FP) és a hamis negatív (FN) képpontok arányát, illetve az IoU-t is. Az eredmények az 1. táblázatban láthatóak, tartalmazva a futási időt is.

A táblázat szerint, a módszerek pontossága alig csökken a mérettől függően, viszont sokkal gyorsabb, mint a többi módszer. A MobileNetV3 kitűnik, azzal, hogy a modell mérete és az FP aránya minimális, de futási időben és pontosságban alul marad. Sőt, a MobileNetV3 esetében az FN aránya nagyon magas, ami azt jelenti hogy nagy valószínűséggel fogunk fontos képpontokat szegmentálni. Ebből a szempontból az FN minimalizálása fontosabb, mint az FP minimalizálása. A Mask R-CNN pontossága hasonló az FPN pontosságához, viszont a modell mérete, a futási idő és az FP aránya a legmagasabb.

A módszerek számszerű összehasonlítása

1. táblázat

	Pontosság [%]	FP [%]	FN [%]	IoU [%]	Idő [mp]
FPN-nagy	94.7	3.36	1.95	77.75	0.022
FPN-közép	94.26	3.08	2.66	76.91	0.018
FPN-kicsi	92.93	4.3	2.77	73.88	0.005
Mask R-CNN	92.71	5.17	2.11	73.16	0.177
MobileNetV3	87.02	2.28	10.7	48.27	0.072

Korábban említettük, hogy a tanításhoz egy Nvidia A100-as kártyát használtunk, az 1. táblázat eredményei is ezen az eszközön futottak le. További kiértékelésként lefuttatuk a saját módszerünk közepes mérettel egyéb eszközökön is. Ezekben a futási idő egy képre a 2. táblázatban látható. Itt megfigyelhető, hogy

a módszerünk elfogadható futási időt ér el beágyazott rendszereken is, mint amilyen egy Jetson Xavier NX, viszont nem ajánljuk a CPU-k használatát.

Futási idő különböző eszközökön

2. táblázat

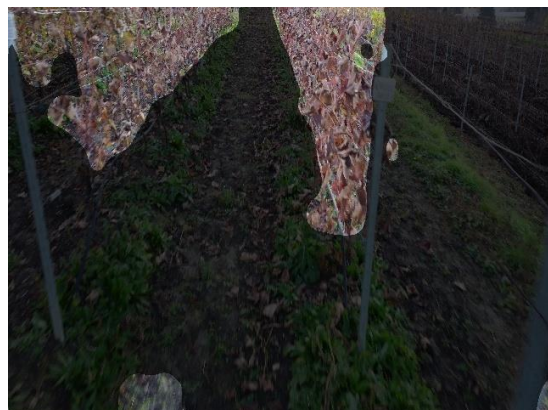
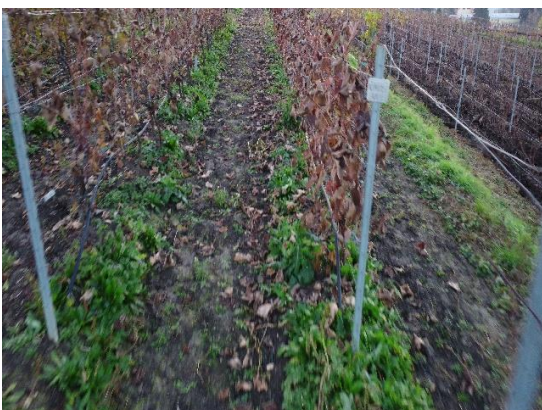
Eszköz	Idő[mp]
Nvidia RTX3080 (10GB)	0.012
Nvidia A100 (40GB)	0.018
Nvidia TeslaT4 (16GB)	0.019
Jetson Xavier NX	0.085
Intel ® Core ™ i9-10900K	0.811
Intel ® Xeon ® Gold 6226R	0.934
Intel ® Core ™ i7-6700K	1.579

A 3. ábrán baloldalon bemutatjuk a módszerünk eredményét, azaz egy kiszegmentált levélzetet, míg a jobb oldalon kiemeljük a valós maszk és a kiszámolt maszk közötti eltéréseket. Ahogy várható volt, a hiba nagy részét a széleknél észleljük.



3. ábra. A szegmentált levélzet (bal), és az eltérések kiemelése (jobb). Kék színnel jelöltük az FP-t, míg pirossal az FN-t.

További tesztek végeztünk szélsőségesebb képeken. Az eredeti adat szettünket koraősszel gyűjtöttük össze, amikor a levelek elkezdtek barnulni. A szélsőséges képeket késő ősszel gyűjtöttük, amikor a levelek már nagyon elbarnultak. Ezt az eredményt láthatjuk a 4. ábrán.



4. ábra. Szélsőséges szegmentálás, késő ősszel. Eredeti kép (bal), kiszegmentált levélzet (jobb).

Egy másik szélsőséges tesztet is végeztünk a késő őszi képekkel. Ezeket a képeket nem a DJI Mini 2-es drónunkal készítettük, hanem egy speciális kamera felszereléssel, amit a [1] dolgozat ír le. Lényegében egy

nagyon erős vakuval készültek a képek, amelyek éjszakai hatást keltenek. A szegmentálás eredménye egy ilyen képre az 5. ábrán látható.



5. ábra. Vakuval készített kép szegmentálása. Eredeti (bal), kiszegmentált (jobb).

4. ÖSSZEFOGLALÁS

Ebben a dolgozatban bemutattunk egy levél szegmentáló módszert, ami hatékonyan működik különböző eszközökön. Az FPN architektúra megfelelő alapot biztosít a megfelelő pontosság eléréséhez úgy, hogy mindeközben a futási idő minimalizálva marad már módszerekhez képest. Habár az elért pontosság ígéretes, további adatokat szeretnénk készíteni, amivel tovább szeretnénk optimalizálni a módszerünk hatékonyságát. Ezen felül a különböző szinterek is pozitívan hatnak az elért eredményekre.

KÖSZÖNETNYILVÁNÍTÁS

A dolgozat készítői hálásak a romániai Analog Devices cégnek, a felajánlott Nvidia eszközökért. Ezt a munkát a Román Nemzeti Tudományos Kutatásokért Felelős Hatóság pénzügyileg támogatta a PN-III-P2-2.1-PED-2021-3120 számú projekt valamint a MTA támogatásáért Domus Alapítvány is.

IRODALMI HIVATKOZÁSOK

- [1] F. Abdelghafour, B. Keresztes, A. Deshayes, C. Germain, J.-P.D. Costa, An annotated image dataset of downy mildew symptoms on Merlot grape variety, *Data Brief.* 37 (2021) 107250.
- [2] D. Aghi, S. Cerrato, V. Mazzia, M. Chiaberge, Deep Semantic Segmentation at the Edge for Autonomous Navigation in Vineyard Rows, in: *IEEEERSJ Int. Conf. Intell. Robots Syst. IROS 2021 Prague Czech Repub. Sept. 27 - Oct. 1 2021*, IEEE, 2021: pp. 3421–3428.
- [3] S. Castriota, *Wine Economics*, The MIT Press, 2020.
- [4] D. Dey, L. Mummert, R. Sukthankar, Classification of Plant Structures from Uncalibrated Image Sequences, in: *IEEE Workshop Appl. Comput. Vis. WACV 2012 Breckenridge CO USA January 9-11 2012*, IEEE Computer Society, 2012: pp. 329–336.
- [5] L. Ghiani, A. Sassu, F. Palumbo, L. Mercenaro, F. Gambella, In-Field Automatic Detection of Grape Bunches under a Totally Uncontrolled Environment, *Sensors.* 21 (2021) 3908.
- [6] S. Gutiérrez, I. Hernández, S. Ceballos, I. Barrio, A.M. Díez-Navajas, J. Tardaguila, Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions, *Comput. Electron. Agric.* 182 (2021) 105991.
- [7] K. He, G. Gkioxari, P. Dollár, R.B. Girshick, Mask R-CNN, in: *IEEE Int. Conf. Comput. Vis. ICCV 2017 Venice Italy Oct. 22-29 2017*, IEEE Computer Society, 2017: pp. 2980–2988.
- [8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature Pyramid Networks for Object Detection, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017: pp. 2117–2125.
- [9] S. Molnár, B. Kelényi, L. Tamás, Feature Pyramid Network Based Efficient Normal Estimation and Filtering for Time-of-Flight Depth Cameras, *Sensors.* 21 (2021) 6257.
- [10] S. Nuske, S. Achar, T. Bates, S. Narasimhan, S. Singh, Yield Estimation in Vineyards by Visual Grape Detection, in: *2011 IEEEERSJ Int. Conf. Intell. Robots Syst. IROS 2011*, pp. 2352–2358.
- [11] K.P. Seng, L.M. Ang, L.M. Schmidtke, S.Y. Rogiers, *Computer Vision and Machine Learning for Viticulture Technology*, *IEEE Access.* 6 (2018) 67494–67510.
- [12] N.D. Tillett, Robotic manipulators in horticulture: a review, *J. Agric. Eng. Res.* 55 (1993) 89–105.