

Gépi látás és mesterséges intelligencia alkalmazása a precíziós növénytermesztésben

1. Bevezetés

A precíziós mezőgazdaság megvalósításához elengedhetetlenek a részletes és pontos monitoring adatok. Ezek lehetnek például pontos meteorológiai információk, a talaj tápanyag-ellátottságára vonatkozó adatok, vagy a növények fejlődését leíró megfigyelések. Ezeket az adatokat mesterséges intelligencia segítségével elemezhetjük, így lehetőség nyílik az optimális beavatkozások meghatározására és végrehajtására.

A monitoringhoz különböző eszközök állnak rendelkezésre, mint például műholdak, műholdfelvételek, drónok és kamerarendszerek. Ezen vizuális adatok feldolgozása és értelmezése a gépi látás feladata. A kutatás középpontjában a gépi látás és a mesterséges intelligencia alkalmazása áll. Egyrészt a paradicsom földközeli kameraképeinek elemzése, hozambecslése, valamint esetleges betegségek korai felismerése. Másrészt a gyomok távérzékeléssel történő detektálása, gyomnyomás térképek előállítására és a gyomok osztályozására kéri.

A hagyományos képfeldolgozási technológiák a képfeldolgozás egyik legklasszikusabb formája. Lényege, hogy a képeken különféle szűréseket

~~¹ Wittmann Antal Növény-, Állat- és Élelmiszer-tudományi Multidiszciplináris Doktori Iskola;~~

² Axiál Kft.;

Témavezető: Dr. Nyéki Anikó Éva egyetemi docens, Albert Kázmér Mosonmagyaróvári Kar Biológiai Rendszerek és Precíziós Technológiai Tanszék;
Vállalati szakértő: Dr. Mesterházi Péter Ákos, Axiál Kft. precíziós gazdálkodási csoportvezető;

alkalmaz annak érdekében, hogy bizonyos jellemzőket kiemeljen. Célja a vizsgált objektumok háttértől való elválasztása. Az objektumok jellemzőit kiemeli, legyen az színelapú információ, textúra- vagy felületjellemző, mint például a Histograms of Oriented Gradient (HOG) (Dalal & Triggs, 2005) jellemzők, illetve különféle alaktényezők. Ezeket jellemzően vektoros formában normalizáljuk, majd különböző gépi tanulási algoritmusok segítségével például Support Vector Machine (SVM) (Mokhtar et al., 2015) vagy a Random Forest algoritmussal (Wójtowicz et al., 2021) elemzik és osztályozzák.

A mesterséges intelligenciára támaszkodó Konvolúciós Neurális Hálózat (CNN)-alapú képfeldolgozási technológiák megközelítése a teljes képet bemenetként használja, és a képen található objektumok jellemzőit próbálja közvetlenül megtanulni. Ehhez egy neurális hálózatot alkalmaz, amely képes különféle összefüggések felismerésére. A tanulási folyamathoz nagyszámú kép és az azokon található célobjektumok pontos megjelölése szükséges. Maga a tanulás a képi adatbázis ismételt feldolgozásával történik, miközben a hálózat a hibák alapján módosítja belső súlyait. Az így betanított hálózat végül képessé válik arra, hogy hasonló, új képeken is felismerje a kívánt célobjektumokat.

Egyik legfontosabb feladat a képi adatbázisok létrehozása. A mesterséges intelligencia alapú megközelítés nagyon sok képet igényel a modellek betanítása során. A képeket nagy hatékonysággal már létre tudjuk hozni, viszont a modell betanításához a képeken lévő cél objektumok címkéjét, koordinátáit, befoglaló kereteit illetve pixel szintű maszkjára is szükség van. Mezőgazdasági képeken jellemzően az objektumok (termések, levelek, virágok) részben vagy egészben kitakarják egymást. Automata képrögztítés és a szél miatt rendszeres a

bemozdult, elmosódott kép. A háttérben lévő többi sorok is gyakran látszanak, így azon látható termések és levelek is gondot okoznak. Ezért nehéz a mezőgazdaságban jó minőségű képi adatbázist létrehozni illetve találni.

A fenti problémák egyik lehetséges megoldása a szintetikus képek létrehozásában rejlik. A szintetikus képek annyit jelentenek, hogy mesterségesen felépített vagy generált képek. Moreno és munkatársai (2023) saját gyom-adatbázist készítettek egy paradicsomültetvényen, ahol kézzel címkézték a képeken látható gyomokat. Minden gyomfajból 30 különböző képet választottak, figyelembe véve a háttér és a fényviszonyok változatosságát. Ezekkel a képekkel egy Stable Diffusion modellt tanítottak be, amely fotórealisztikus képeket generált. Az új képeket zöld maszkkal és automatikus befoglaló doboz címkéssel egészítették ki, majd hozzáadták a tanító adathalmazhoz. A kibővített adathalmazon tanított CNN modell jobb eredményeket ért el, mint az eredeti kisméretű adathalmazon. Az mAP érték vegyes adathalmaz esetén 6–9%-kal javult.

Sahu és munkatársai (2023) a Deep Dream neurális hálózatot használták szintetikus képek létrehozására, a Plant Village adatbázis alapján. A módszer felerősítette a levélbetegségek vizuális jegyeit. Az így keletkezett képekből CNN-alapú módszerrel jellemzőket nyertek ki, melyeket hagyományos gépi tanulási algoritmusokkal (RF, AdaBoost, SGB) osztályoztak. Az eredmények azt mutatták, hogy a szintetikus képek javították az osztályozási pontosságot.

Upreti és munkatársai (2024) generatív ellenséges hálózatot (GAN) használtak az eredeti képek bővítésére és az osztályok közötti egyensúly megteremtésére. Saraswathi és Banu (2024) a Plant Village adatbázist

egészítették ki Hybrid CGAN segítségével előállított szintetikus képekkel, majd hagyományos képfeldolgozási módszerekkel vektorokat vontak ki, melyeket kNN és SVM osztályozókhoz használtak fel.

2. Anyag és Módszer

2.1. Paradicsom detektálása

A paradicsomnövények felismerése érdekében egy minta területen 155 tő paradicsomot monitoroztunk rendszeresen. A paradicsomok 12 sorban voltak ültetve és QR kódokat helyeztünk ki a tövekhez, hogy minden tő beazonosítható legyen. A felvételezés nyolc alkalommal történt, körülbelül kéthetes időközönként. Minden egyes növényről négy különböző irányból készítettünk képeket. Nagy mennyiségű képi adatbázist gyűjtöttünk, ahol egy-egy alkalommal összesen 620 kép készült, ami összesen 4960 felvételt jelentett.

A felvételek 3120x4160 pixeles felbontásba készültek, egy állványról, amely 1 méterre helyeztünk el a tövektől és a kamera 1 méter magasan helyezkedett el. A képeket rendszereztük, időpont-, irány -, pozíció-, és sorok szerint.

A kutatás első szakaszában a piros paradicsomok felismerésére volt. A képeken teljesen látható, részben takart, valamint klaszterekben elhelyezkedő paradicsomokat azonosítottunk, amelynek eredményeként 656 egyedi mintát nyertünk ki. Ezek a mintaadatok önmagukban nem bizonyultak elegendőnek egy CNN hatékony tanításához, ezért szintetikus képeket generáltunk tréningadatoknak. Ennek keretében 2600 tanítókép készült, körülbelül 26 000

paradicsom pozícióval, továbbá 800 validációs kép (nagyjából 8000 paradicsommal) és 520 tesztkép (mintegy 5000 paradicsommal). Az így előállított képi adathalmazon egy You Only Look Once (YOLO)v11 nano (Khanam & Hussain, 2024) CNN hálózatot tanítottunk be.

A másik eljárás a tövek szegmentálása volt a háttértől. A képeken a háttérben gyakran más paradicsomsorok növényei is megjelennek, ezért az elemzést megelőzően fontos ezek kiszűrése. Ezeknek a problémáknak a kezelésére egy mély neurális hálózatot, a Depth Network modellt alkalmaztuk (Kumar et al., 2018), amely mélységi képek előállításával segítette a paradicsomok hatékony szegmentálását. A kapott mélységi információk alapján egy közel-távol maszkot hoztunk létre, amely segítségével hatékonyan szegmentáltuk a vizsgálni kívánt paradicsomtöveket.

2.2. Közeli növények osztályozása

Egy korábbi kutatás során 500-500 képből álló adathalmazt alkalmaztunk hat különböző növényfaj osztályozására, amely során 94%-os pontosságot értünk el (Moldvai et al., 2024).

Ezek a képek egy nyilvánosan elérhető adatbázisból származnak, amely nem kereskedelmi célú alkalmazásokhoz hozzáférhető a következő címen: <https://github.com/zhangchuanyin/weed-datasets>. A képek felülről lefelé készültek kb. 30 cm-re a növénytől.

A képek előfeldolgozása során RGB színtérből HSV színtérré alakítjuk. A zöld részeket vagyis a növényeket kiemeljük és külön osztályokba tesszük. A legnagyobb felületű növényt meghagyjuk a többit kiszűrjük a képről. Optimális

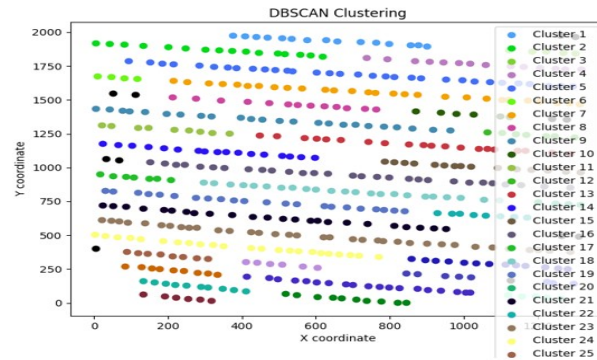
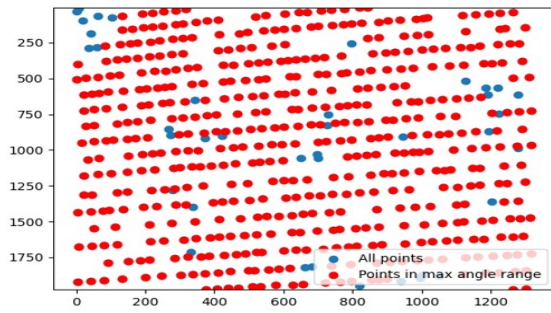
esetben ez egy növényt jelent az adatbázis jellegzetessége miatt. Sajnos néha több növény összeér, belelóg, ami miatt az összes létrehozott maszkolható képet át kell nézni és manuálisan javítani a problémás képeket. Így jön létre egy szintetikus képek cél objektumait tartalmazó adatbázis.

A második lépésben 24 446 háttérkép készült 1024x1024 pixeles méretben. A háttérképek és cél objektum adatbázis képeiből, szintetikus képeket állítottunk elő. A háttérképre véletlenszerűen 2 és 20 közötti számú növényt maszkoltunk rá. Az így generált tréningadatok száma 4900 kép lett, ami körülbelül 80 000 növényt foglalt magában. A validáláshoz 1400 képet használtunk fel (nagyjából 20 000 növénnel), míg a teszteléshez 640 képet (hozzávetőleg 10 000 növénnel).

Ezen képekkel egy YOLOv11 nano rendszert tanítottunk be, majd teszteltük szintetikus képeken és az valós képeken is. A szintetikus képeken a Precision, Recall, mAP50 és mAP50-95 metrikákat számoltuk ki.

2.3. Távérzékelte adatok elemzése

A kutatás másik fő területe a gyomnövények azonosítása volt távérzékelte adatokon kapás kultúrákban. Ehhez drónfelvételeket használtuk 21.7 méter magasságból felvételezve 5280x3956 pixel méretben. A felvételek jellemzően napraforgó- és kukoricatáblákról származtak.



1. ábra: A pontok szétválasztása kultúrnövény és gyomnövényre illetve a sorok klaszterbe szervezése.

A képfeldolgozás első lépéseként az RGB képet HSV színtérbe konvertáltuk, majd a Hue csatorna alapján zöld maszkot hoztunk létre. A zöld foltok koordinátáiból (X, Y) pontadatbázist állítottunk össze, amelyen szomszédossági vizsgálatot végeztünk, és távolság- és szöghisztogramokat számoltunk. A hisztogram maximumai alapján meghatároztuk a kultúrnövény sorainak irányát és a jellemző tőtávolságot. Ezek alapján azonosítottuk a sorba eső pontokat, majd a pontokat DBSCAN algoritmussal klaszterekbe rendeztük (1. ábra), amelyekre egyeneseket illesztettünk. A pontokat ezután sorban lévő és sorok között elhelyezkedő pontokra osztályoztuk. Csak a sorok közti pontokat felhasználva készítettük el a gyomnyomás térképet, amelyből egy előre meghatározott küszöbérték alapján kijuttatási térképet generáltunk.

3. Eredmények

3.1. Paradicsom bogyó detektálásának eredményei

Az elemzés során nehézséget okozott, hogy a háttérben további paradicsomtövek is megjelentek, valamint a sárga virágok és a közeli napraforgók vizuálisan zavarták a képek értelmezését (2. ábra).



2. ábra: a) eredeti kép; b) mélységi kép; c) távoli sorok kiszűrve;

Az eredmények arra utalnak, hogy a módszer megfelelő irányt képvisel. A 2. b) ábrán látható, hogy mélységi kép létrejött és sikeresen kiszűrtük a lépről a hátsó sorok zavaró részleteit 2. c) ábra. Így CNN-nel elemezve már nem okoznak problémát. A módszert még finomítani szükséges ezért további kiértékelések fognak következni.

A betanított YOLOv11 nano érett paradicsom felismerési képességét a 3. ábra szemlélteti. A szintetikus adatokon a modell (Precision 92,96%; Recall 93,85%; mAP50 97,50%; mAP50-95 95,84%) minőségi mutatókat ért el.



3. ábra: Detektál érett paradicsomok YOLOv11nano modell betanítása után.

3.2. Közeli növények osztályozásának eredményei

Az adathalmazon egy YOLOv11nano CNN hálózatot tanítottunk be. A szintetikus képi teszt adatokon a modell (Precision 97,44%; Recall 97,47%; mAP50 98,59%; mAP50-95 94,10% minőségi mutatókat ért el. Az valós képeken (4. ábra) a felismerés és lokalizáció megfelelőnek bizonyult, azonban néhány hibás detektálás is előfordult. A rendszer további finomhangolása és a felismerni kívánt osztályok bővítése jövőbeli feladatok közé tartozik.

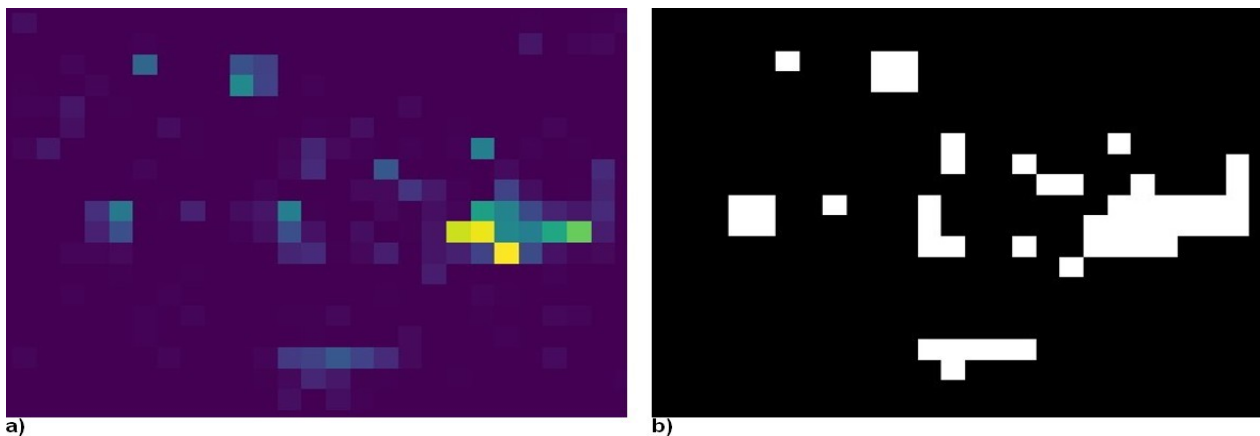


4. ábra: Detektálás eredménye valós képeken.

3.3. Távérzékelte adatok elemzése, gyomnyomás térkép és kijutatási térkép

A kutatás során sikeresen elvégeztük a növényi sorok és sorközök detektálását. A kultúrnövények klaszterezése és az egyenesek illesztése a sorokra megtörtént. A kultúrnövények és a gyomnövények osztályozása elhelyezkedés alapján megtörtént.

A gyomnövények gyakorisága alapján úgynevezett heatmapet állítottunk elő (5.a ábra). Majd egy küszöb megadása után két állapotú kijutatási térképet hoztunk létre (5.b ábra).



5. ábra: a) gyom gyakoriság térkép; b) kijutatási térkép;

4. Összegzés

Az elmúlt időszakban végzett kutatás során jelentős előrelépéseket értünk el mind a paradicsom detektálás, mind a gyomfelismerés terén. Több nagy képi adatbázist építettünk, melyek feldolgozását elkezdtük. A képek feldolgozása során a hagyományos gépi látás módszereire és gépi tanulás algoritmusaira is támaszkodtunk, de a CNN alapú Depth Network és a YOLOv11nano mesterséges intelligencia alapú rendszereket is alkalmaztuk. A jövőbeni munka

a modellek további finomítására, az adathalmazok bővítésére és az algoritmusok robusztusságának növelésére irányul.

Az elmúlt 2 évben elterjedt Transformer alapú neurális hálózatok - amelyek első sorban szövegek értelmezésére és szöveggenerálásra használtak - egyik változata képes képi adatokat is elemezni. Ezeket hívják Vision Transformersnek (ViT). A ViT a képet globálisan nézi szembe a CNN alapú rendszerekkel, amelyek lokálisan keresi a mintákat. Mezőgazdasági képek elemzésére is nagyon ígéretes technológia, habár nagyságrenddel több tanító mintára van szüksége, mint a CNN rendszereknek. Erre viszont a fent bemutatott szintetikus képek adhatnak jó megoldást.

A következő időszak fő feladatának a már bemutatott módszerek javítását és a ViT alkalmazását, eredményeinek a CNN rendszerek eredményeivel való összehasonlítását tervezem.

Felhasznált irodalom

- Dalal, N., & Triggs, B. (2005). (HOG) Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 05)*, 1, 886 - 893. <https://doi.org/10.1109/CVPR.2005.177>
- Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2410.17725>
- Kumar, A. C. S., Bhandarkar, S. M., & Prasad, M. (2018). DepthNet: A Recurrent Neural Network Architecture for Monocular Depth Prediction. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 396 - 3968. <https://doi.org/10.1109/CVPRW.2018.00066>
- Mokhtar, U., El Bendary, N., Hassenian, A. E., Emary, E., Mahmoud, M. A., Hefny, H., & Tolba, M. F. (2015). (SVM) -Based Detection of Tomato Leaves Diseases. In D. Filev, J. Jabłkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szykarczyk, & S. Zadrozny (Eds.), *Intelligent Systems' 2014* (Vol. 323, pp. 641 - 652). Springer International Publishing. https://doi.org/10.1007/978-3-319-11310-4_55

- Moldvai, L., Mesterházi, P. Á., Teschner, G., & Nyéki, A. (2024). Weed Detection and Classification with Computer Vision Using a Limited Image Dataset. *Applied Sciences*, *14*(11), 4839. <https://doi.org/10.3390/app14114839>
- Moreno, H., Gómez, A., Altares-López, S., Ribeiro, A., & Andújar, D. (2023). Analysis of Stable Diffusion-derived fake weeds performance for training Convolutional Neural Networks. *Computers and Electronics in Agriculture*, *214*, 108324. <https://doi.org/10.1016/j.compag.2023.108324>
- Sahu, P., Chug, A., Singh, A. P., & Singh, D. (2023). Classification of crop leaf diseases using image to image translation with deep-dream. *Multimedia Tools and Applications*, *82*(23), 35585 - 35619. <https://doi.org/10.1007/s11042-023-14994-x>
- Saraswathi, E., & Banu, J. F. (2024). Hybrid CGAN-based plant leaf disease classification using OTSU and surf feature extraction. *Neural Computing and Applications*, *36*(23), 14395 - 14407. <https://doi.org/10.1007/s00521-024-09812-w>
- Upreti, K., Singh, P., Jain, D., Pandey, A. K., Gupta, A., Singh, H. R., Srivastava, S. K., & Prasad, J. S. (2024). Progressive loss-aware fine-tuning stepwise learning with GAN augmentation for rice plant disease detection. *Multimedia Tools and Applications*, *83*(37), 84565 - 84588. <https://doi.org/10.1007/s11042-024-19255-z>
- Wójtowicz, A., Piekarczyk, J., Czernecki, B., & Ratajkiewicz, H. (2021). (RF) A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. *Journal of Photochemistry and Photobiology B: Biology*, *223*, 112278. <https://doi.org/10.1016/j.jphotobiol.2021.112278>
-