

Doktori (PhD) értekezés tézisei

Gépi és mély tanulási eljárások alkalmazása a városi környezet vizsgálatában, nagyfelbontású és különböző dimenziójú távérzékelt adatok alapján

Abriha Dávid

Témavezető:

Prof. Dr. Szabó Szilárd, egyetemi tanár



DEBRECENI EGYETEM

Földtudományok Doktori Iskola

Debrecen, 2023

Bevezetés

Az elmúlt évtizedekben növekvő tendenciát mutató urbanizáció hatására a természetes környezet egyre jelentősebb mértékben alakul át. A globális népességnövekedés és a megavárosok (10 millió főt meghaladó lakosságszám) gyarapodó száma egyre nagyobb nyomást gyakorolnak a Föld ökoszisztémáira. Az ilyen léptékű változások számos gazdasági, társadalmi és környezeti kihívásokat generálnak, amelyeknek a kezelése egyre több városban kap kiemelt prioritást, és ennek eredményeként a fenntarthatóság és a fenntartható urbanizáció lett a városfejlesztési stratégiák egyik legfontosabb alapelve (Beaverstock et al., 1999; Kadhim et al., 2016; Shen et al., 2012). Doktori értekezésem fókuszában az épített környezet, azon belül is az épületek nagyfelbontású és különböző dimenziójú távérzékelte adatokkal végzett elemzése állt, elsődlegesen Debrecen városában.

Bár a városi régiók rendkívül összetettek, a távérzékelés széles körben kínál lehetőségeket a döntéshozóknak ezen kihívások leküzdéséhez. Kutatásom során kiemelt szerepet kapott a tradicionális machine learning (gépi tanulás) módszerekkel végzett képosztályzás, illetve az ehhez szorosan kapcsolódó pontosságellenőrzés, valamint napjaink legkorszerűbb technológiájával, deep learning (mély tanulás) alapú megközelítéssel végzett épületszegmentáció is.

Munkám során a céloom elsőként különböző tetőfedő anyagok osztályozhatóságának a vizsgálata volt, eltérő körülmények között. Egy kiválasztott mintaterületen a három leggyakrabban előforduló tetőfedő típust vontam be az osztályozásba, melyet WorldView-2 műholdfelvétel alapján végeztem el. Következő lépésként megvizsgáltam, hogy az osztályozások során a területi autokorreláció milyen hatást gyakorol a modellek pontosságára. A tetőfedő anyagok osztályozása előtt célszerű egy olyan maszk előállítását és a felvételen történő alkalmazását, amely csak ezeket a célobjektumokat tartalmazza, ezzel csökkentve a szükséges feldolgozási kapacitást. A kutatásom során szintén céloom volt megvizsgálni, hogy a relatívan olcsón előállítható vagy beszerezhető felvételeken végzett, modern, mély konvolúciós neurális hálózatokon alapuló objektum detektálás milyen pontossággal alkalmazható épületek kinyerésére.

Célkitűzések

Kutatásaim során az alábbi kérdésekre kerestem a választ:

1. Hogyan befolyásolja a WorldView-2 műholdfelvételen végzett tetőfedő anyagok osztályozását a térbeli felbontást javító pansharpening eljárás, valamint milyen hatást gyakorol a tematikus pontosságra az egyes tetőfelületek napsütötte és árnyékos kategóriákra történő szétválogatása?
2. Milyen mértékben járul hozzá a magas pozitív területi autokorreláció az osztályozási pontosság túlbecsléséhez, továbbá hogyan lehet kiküszöbölni a területi autokorreláció torzító hatását a k-szoros keresztvalidációval történő pontosságbecslés megbízhatóságának biztosítása érdekében?
3. Mennyire hatékony a U-Net konvolúciós neurális hálózat a különböző forrásból származó multispektrális távérzékelte adatokon alapuló épületszegmentálásban, kevés tanító adat felhasználásával?

Anyag és módszer

Képosztályozás és a területi autokorreláció vizsgálata

A tetőfedő anyagok alapján végzett klasszifikációs vizsgálatokat és az autokorrelációhoz kapcsolódó pontossági elemzéseket 2 légi ortofotó (debreceni és vaihingeni mintaterületek) és 1 műholdfelvétel (WorldView-2; WV2) felhasználásával végeztem el.

A debreceni mintaterület esetében a képosztályozáshoz szükséges referenciaadatokat terepi bejárások alkalmával gyűjtöttem, ezek során 350 ház tetőfedő anyagának típusát azonosítottam. A területen a 3 leggyakoribb tetőfedő alapján végeztem el az osztályozásokat, amelyek a vörös cserepek, a sötét cserepek (fekete, szürke, barna cserepek), illetve az azbeszt-cement tartalmú panelek voltak. Bár az adatgyűjtés alatt egyéb anyagokat is azonosítottam (fém, kátrány stb.), ezeket az alacsony előfordulási gyakoriságuk miatt kihagytam az elemzésekből. A németországi (Vaihingen) mintaterület esetében a Google Earth és a Google Street View felvételeit használtam a különböző tetők azonosítására. Mivel itt azbeszt tetők nem voltak jelen, azonban lapostetőt nagy mennyiségben tudtam

azonosítani, így a sötét és vörös cserép mellett ez lett a harmadik kategória.

A képosztályozások során, amikor a pansharpening eljárásnak, valamint a tetőfelületek megvilágítottság alapján történő szétválogatásának tematikus pontosságra gyakorolt hatását vizsgáltam, a klasszifikációkat kétféle megközelítés alapján végeztem el. Alkalmaztam egyrészt hagyományos, többváltozós statisztikai osztályozási módszereket (lineáris diszkriminancia-analízis [LDFA]; kvadratikus diszkriminancia-analízis [QDFA]), valamint robusztus gépi tanulási algoritmust (*Random Forest*; RF) is.

A távérzékelte felvételekről elmondható, hogy azok területi autokorrelációval (*spatial autocorrelation*; SA) erősen terheltek, mivel a szomszédos pixelek értékei egymáshoz nagyon hasonlóak (Wulder és Boots, 1998). Bár a távérzékelte adatoknak ez egy általános jellemzője, azonban ez egyúttal megkérdőjelezheti a keresztvalidációval kapott eredmények hitelességét amennyiben a referenciaadatok kiválasztása nem megfelelő.

Az SA pontosságra gyakorolt torzító hatásának számszerűsítése érdekében az egyes modellek teljesítményének meghatározását kétféleképpen végeztem el: (i) a rétegzett k -szoros keresztvalidáció (*stratified k-fold cross validation*; SCV) alapján számított becslött pontosság, és (ii) a független tesztadatokkal történő validálással kapott tényleges pontosság alapján.

Annak érdekében, hogy a keresztvalidáció alkalmas legyen a modellek tényleges pontosságának becslésére, Python-ban fejlesztettem egy módszert, amely a referenciaadatokat azok térbeli elhelyezkedése alapján szelektálja: csak a paraméterként meghatározott távolságon kívül eső pontok kerülnek kiválasztásra a modellezéshez.

Az SA pontosságra gyakorolt torzításának mértékét összesen 7 különböző mintavételezési módszer esetén vizsgáltam. A validáció során a pontossági értékek minden esetben 50 futtatás alapján kerültek meghatározásra, kivéve a kontroll mérésnek számító, tényleges pontosság céljából végzett független tesztelést.

Épületszegmentáció mély konvolúciós neurális hálózat alapján

Kutatásom során a U-Net (Ronneberger et al., 2015) mély tanulási architektúrával végzett épületszegmentáció hatékonyságát 2

műholdfelvétel (WV2, WV3) és 1 légi ortofotó alapján vizsgáltam, debreceni mintaterületen.

A U-Net konvolúciós neurális hálózattal történő modellezés kétféle input adattal történik: egyrészt szükség van az eredeti raszterre, valamint az ehhez tartozó maszkra, mely utóbbi az esetemben egy bináris maszkot jelent, amely az épületeket különíti el az összes többi objektumtól. A modellezés előtt az input adatokat $512 * 512$ pixel méretű darabokra vágtam. Az ortofotó és a WV2 felvétel esetében egy nagyjából 2 km^2 -es területen került vektorizálásra az összes ház, míg a WV3-nál ez a terület $0,3 \text{ km}^2$ volt. A két terület közötti eltérés oka, hogy a WV3 esetében a kevesebb tanító adattal történő modellezés értékelése is részét képezte a kutatásnak.

A tanító adatok manuális gyűjtése (kézi vektorizálása) nagyon időigényes feladat. A kutatás során a 3 felvételen összesen több, mint 3000 épület lett vektorizálva. Bár az U-Net architektúra alapvetően jól működik viszonylag kis számú adattal is, az augmentáció (adatbővítés) elengedhetetlen lépés a jobb tematikus pontosság elérése érdekében. A kutatásom során a tradicionálisnak tekinthető augmentációs módszereket alkalmaztam: az eredeti képeket vízszintesen és függőlegesen tükröztem, valamint 90, 180 és 270 fokkal elforgattam azokat. Így az eredeti kép mellett az augmentációnak köszönhetően 5 további képet kaptam.

Az egyes modelleket legfeljebb 50 *epoch*-on (a tanítási iterációk száma) keresztül tanítottam. Implementáltam egy olyan, korai leállító mechanizmust, amely megszakította a tanítási folyamatot, ha a modell teljesítménye nem javult 15 *epoch* alatt. Minden modell esetében 0,001-es kezdeti tanulási rátát (*learning rate*) választottam. Optimalizációs algoritmusként az Adam-módszert használtam, 20-as *batch*-mérettel. Összesen 40 „eredeti”, azaz augmentáció előtti képet használtam a WV2 és a légi ortofotó esetében, és 5, valamint 10-et a WV3 esetében. A modellezés Python programozási nyelven történt a TensorFlow (Abadi et al. 2016) csomag segítségével. Minden modellfuttatáshoz 24 GB VRAM-mal rendelkező Nvidia RTX 3090 videokártyát használtam.

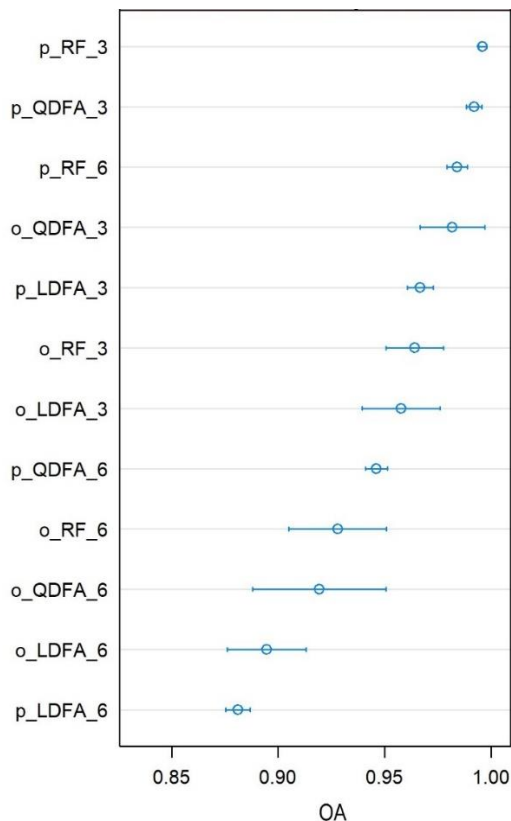
Kutatási eredmények

1. tézis

WorldView-2 műholdfelvételen végzett tetőfedő anyagok osztályozása kapcsán kimutattam, hogy az egyes kategóriáknak a megvilágítottságuk alapján történő felosztása a tematikus pontosság csökkenését, a pansharpening eljárással végzett térbeli felbontás javítása pedig a növelését eredményezi.

A WV2 felvételen végzett klasszifikációkhoz 3 kategóriát jelöltem ki (azbeszt, vörös cserép, sötét cserép), melyeket 3 algoritmus alapján osztályoztam (LDFA, QDFA, RF). Továbbá a vizsgálatokat egyrészt elvégeztem a tetőfedő anyagok kategóriáinak összevont, valamint napsütötte és árnyékos területekre történő szétválogatása alapján is, így összesen 12 modellt elemeztem (*1. ábra*).

Megállapítható, hogy a kategóriáknak a megvilágítottságuk alapján történő felosztása a pontosságok csökkenését eredményezte: OA értékek 6-7%-kal alacsonyabbak voltak abban az esetben, amikor a napsütötte és árnyékos pixelek külön osztályt alkottak. A pansharpening eljárás hatékony technikáknak bizonyult a tematikus pontosságok növelésére: az OA értékek rendszerint 2-3%-os növekedést mutattak, de előfordult olyan eset is (6 kategóriás RF), amikor az eredeti és feljavított modellváltozatok között 6%-os különbség volt megfigyelhető.



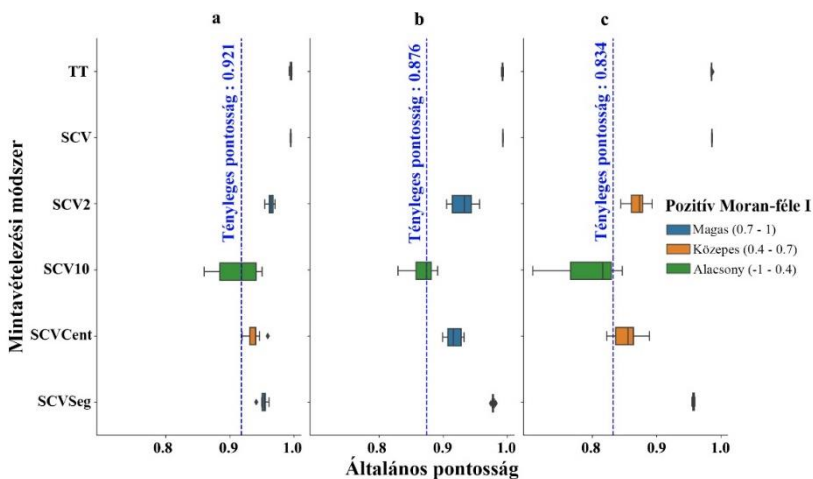
1. ábra. Az egyes modellváltozatok az általános pontosságuk (OA) alapján rangsorolva. o = eredeti felvétel; p = pansharpening eljárással feljavított felvétel; LDFA = lineáris diszkriminancia-analízis; QDFA = kvadratikus diszkriminancia-analízis; RF = Random Forest; 3 vagy 6 = az osztályok száma

2. tézis

Megállapítottam, hogy képosztályozás során a referenciaadatokban fellépő magas pozitív területi autokorreláció nagy mértékben torzítja az osztályozási pontosságokat, amely az adatok térbeli elhelyezkedésük alapján történő újra mintavételezésével hatékonyan megszüntethető.

A magas SA következtében, a pontosságvizsgálat során fellépő torzító hatás nagyon jelentős volt azokban az esetekben, amikor a referenciaadatokat véletlenszerűen osztottam fel tanító és tesztelő területekké, valamint, ha a napjainkban egyre népszerűbbé váló SCV módszert alkalmaztam a pontosságok becslésére: a becült pontosságok és a független tesztadatokkal végzett tényleges pontossági (TP) értékek közötti különbségek elérték a 17%-ot, így a modellek 99% feletti OA-t eredményeztek (2. ábra).

Az általam Pythonban kidolgozott módszer, azaz a referenciaadatok térbeli elhelyezkedésük alapján történő szelekciója sikeresen megszüntette az említett torzító hatást: a modellek SCV-vel becült pontossága, és a független tesztadatokkal kapott TP értékek jól közelítették egymást. A kisebb térbeli felbontású WV2 (50 cm) felvétel 5-9%-kal teljesített jobban, mint a két ortofotó (9 és 10 cm felbontás), melynek legfőbb oka, hogy a WV2 esetében a tetőt reprezentáló képpontok száma jóval kisebb volt, ezáltal az osztályozáshoz használt kategóriákon belüli pixelértékek szórása is csökkent.



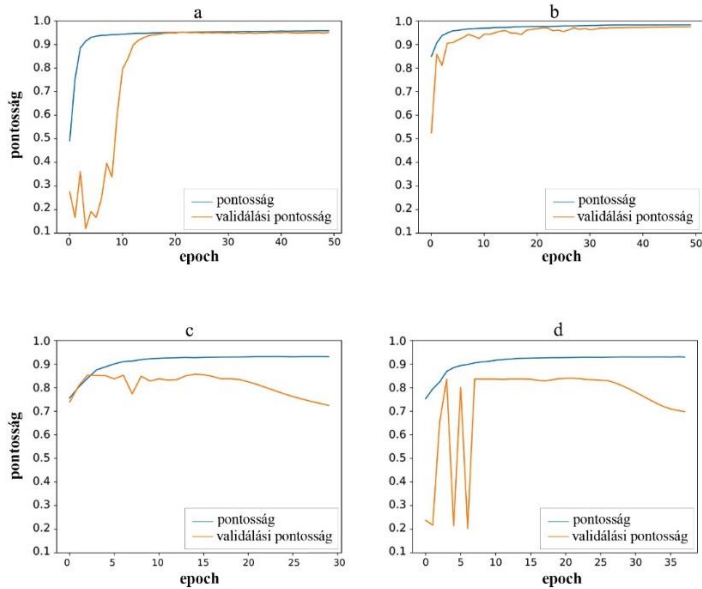
2. ábra. Az egyes mintavételezési módszerek alapján végzett osztályozások eredményei 50 modell alapján, a Moran-féle I vonatkozásában (a = WorldView-2; b = Debreceni ortofotó; c = Vaihingeni ortofotó)

3. tézis

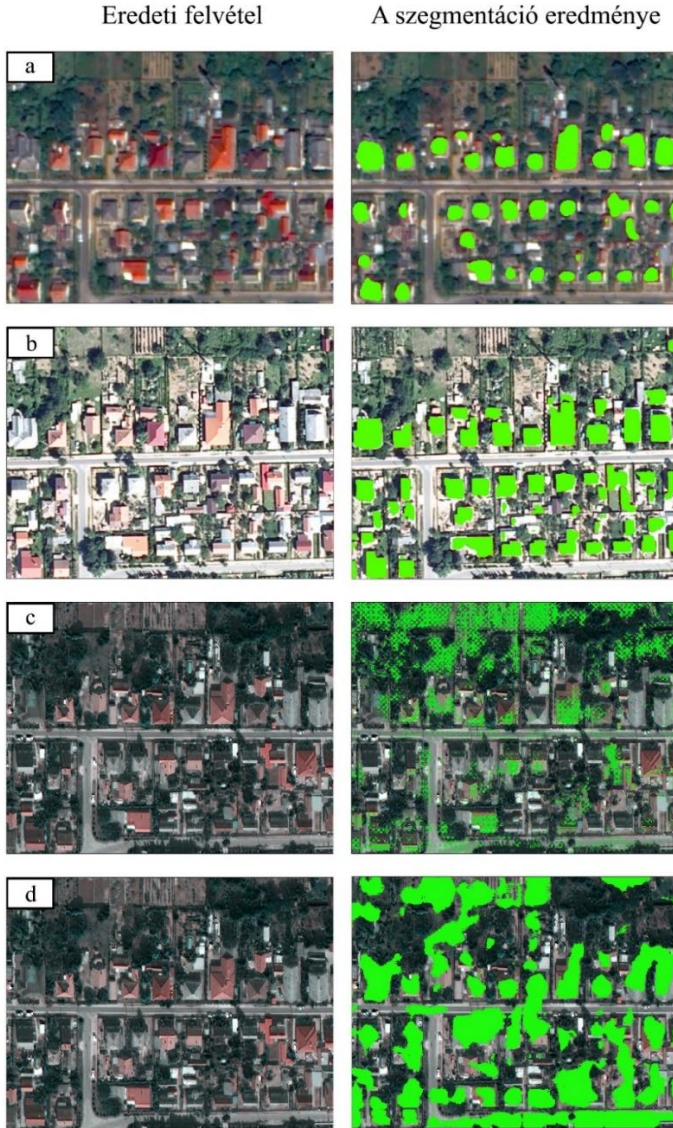
Megállapítottam, hogy a U-Net mély konvolúciós architektúra alacsony számú tanító adat mellett is nagy pontossággal alkalmazható épületszegmentálási feladatokra.

Amennyiben a tanításhoz használt, és a szegmentálni kívánt adatok forrása – azaz a szenzor – megegyezett, a validálási pontosságok már kis mennyiségű tanító adat (192 darab, 512 x 512 pixeles kivágat) esetén is 94% felettiek voltak (2-P40-2, O-P40-O; 3. ábra). A WV3 esetében a tanító adatok további csökkentése (3-P5-3; 3-P10-3; 24 és 48 darab kivágat) a tanulási görbék túlilleszkedését eredményezték.

Amennyiben a validálási pontosság magas volt, és a modell nem illeszkedett túl, a szegmentálás is sikeres volt (WV2, ortofotó; 4/a. és b. ábra): szinte minden épület azonosításra került, és csak néhány helyen jelent meg tévesen szegmentált pixel. Azok az épületek, amelyeket a modell tévesen háttérként (azaz nem épületként) azonosított (fals negatív eredmények), szinte kizárólag kisebb melléképületek (garázsok, fészerek stb.) voltak. A WV3 felvétel alapján betanított két modell esetében (3-P5-3, 3-P10-3) gyenge szegmentálási eredményeket kaptam. Amikor csak 24 WV3 tanító képet használtam a modellezéshez (4/c. ábra), a szegmentálás nem eredményezett összefüggő területeket, és az épületek nagy része kihagyásra került a végeredményből. Bár a 48 tanító képpel végzett elemzés során már néhány épület összefüggő foltokként jelent meg (4/d. ábra), a szegmentálás összességében még mindig gyenge volt, és sok hibát tartalmazott: a csupasz talajfelszín és az utak nagy része tévesen épületként került azonosításra (fals pozitív eredmények).



3. ábra. Tanulási görbék abban az esetben, amikor a tanító adatok a szegmentált felvételtől származtak. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)



4. ábra. A szegmentáció eredménye abban az esetben, amikor a tanító adatok a szegmentált felvételtől származtak. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

4. tézis

Kimutattam, hogy eltérő forrásból származó távérzékelte felvételek fúziója során is nagy pontosságok érhetőek el épületszegmentálás során a U-Net módszerrel, amennyiben legalább kis mennyiségben minden felvételtől felhasználásra kerül tanító adat a modellezés során.

A WV3 felvétel szegmentálása abban az esetben is gyenge eredményt adott, ha a modellezéshez csak a másik két felvétel (WV2, ortofotó) tanító adatait használtam (2O-P80-3): bár a betanítás során a felhasznált kivágatok száma jelentősen nőtt (384 darab), az F1-érték mindössze 0,184 volt. Az előző modellt mindössze 24, a WV3-ból származó kivágattal bővítve (23O-P85-3) a pontosság is jelentősen emelkedett: a szegmentáció eredménye vizuálisan is sokat javult, az F1-érték pedig 0,661-re nőtt. További 24 kivágat bevonása (23O-P90-3; 5. ábra) bár tovább emelte a pontosságot (F1-érték: 0,693; 1. táblázat), ennek mértéke már jóval kisebb volt. Bár itt is előfordultak téves kategorizálások, ezek aránya jelentősen csökkent, és az épületek többségét helyesen azonosította a modell.



5. ábra. A WV3-on végzett szegmentáció eredménye az összes rendelkezésre álló tanító adat felhasználásával (23O-P90-3). – a) eredeti felvétel; b) a szegmentálás eredménye. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

1. táblázat. A modellek pontossága a WV3-on végzett szegmentáció esetében a WV2, az ortofotó és a WV3 tanító adatainak a felhasználásával. Az első és utolsó tag: 2 = WorldView-2; 3 = WorldView-3; O = ortofotó. Második tag = a predikcióhoz használt tanító képek száma (augmentáció előtt)

Modell	Precizitás	Szenzitivitás	F1-érték
3-P5-3	0,22	0,20	0,213
3-P10-3	0,25	0,72	0,369
23O-P85-3	0,69	0,64	0,661
23O-P90-3	0,68	0,71	0,693

Irodalomjegyzék

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv:1603.04467.
- Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities*, 16(6), 445–458. [https://doi.org/Doi10.1016/S0264-2751\(99\)00042-6](https://doi.org/Doi10.1016/S0264-2751(99)00042-6), NASA SP-351, NASA, Washington, D.C., pp. 309-317.
- Kadhim, N., Mourshed, M., & Bray, M. (2016). Advances in remote sensing applications for urban sustainability. *Euro-Mediterranean Journal for Environmental Integration*, 1(1). <https://doi.org/10.1007/s41207-016-0007-4>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Shen, L. Y., Peng, Y., Zhang, X. L., & Wu, Y. Z. (2012). An alternative model for evaluating sustainable urbanization. *Cities*, 29(1), 32–39. <https://doi.org/10.1016/j.cities.2011.06.008>
- Wulder, M., & Boots, B. (1998). Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, 19(11), 2223–2231. <https://doi.org/Doi10.1080/014311698214983>



Nyilvántartási szám: DEENK/158/2023.PL
Tárgy: PhD Publikációs Lista

Jelölt: Abriha Dávid
Doktori Iskola: Földtudományok Doktori Iskola
MTMT azonosító: 10063695

A PhD értekezés alapjául szolgáló közlemények

Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

1. **Abriha, D.**, Kovács, Z., Ninsawat, S., Bertalan, L., Balázs, B., Szabó, S.: Identification of roofing materials with Discriminant Function Analysis and Random Forest classifiers on pan-sharpened WorldView-2 imagery - a comparison.
HunGeoBull. 67 (4), 375-392, 2018. ISSN: 2064-5031.
DOI: <http://dx.doi.org/10.15201/hungeobull.67.4.6>

Idegen nyelvű tudományos közlemények külföldi folyóiratban (2)

2. **Abriha, D.**, Srivastava, P. K., Szabó, S.: Smaller is better? Unduly nice accuracy assessments in roof detection using remote sensing data with machine learning and k-fold cross-validation.
Heliyon. 9 (3), 1-17, 2023. ISSN: 2405-8440.
DOI: <http://dx.doi.org/10.1016/j.heliyon.2023.e14045>
IF: 3.776 (2021)
3. **Abriha, D.**, Szabó, S.: Strategies in training deep learning models to extract building from multisource images with small training sample sizes.
Int. J. Digit. Earth. 16 (1), 1707-1724, 2023. ISSN: 1753-8947.
DOI: <http://dx.doi.org/10.1080/17538947.2023.2210312>
IF: 4.606 (2021)

Magyar nyelvű konferencia közlemények (4)

4. **Abriha, D.**, Szabó, S., Enyedi, P.: Városi objektum kinyerést célzó deep learning algoritmus alkalmazása nagy felbontású légifelvételek alapján.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XII.. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi kiadó, Debrecen, 9-16, 2021. ISBN: 9789633189771
5. **Abriha, D.**, Szabó, S.: A képosztályozás során fellépő területi autokorreláció vizsgálata Python programozási környezetben.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XI. : theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 17-21, 2020. ISBN: 9789633188866





6. **Abriha, D.**, Banka, F., Szabó, S.: Random Forest osztályozó algoritmus pontosságának vizsgálata tetőfedő anyagok azonosításában multispektrális adatokkal.
In: Az elmélet és a gyakorlat találkozása a térinformatikában X.: Theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 19-24, 2019. ISBN: 9789633180549
7. **Abriha, D.**, Szabó, S.: Hiperspektrális távérzékelés alkalmazása tetőfedő anyagok azonosításában.
In: Az elmélet és gyakorlat találkozása a térinformatikában IX.. Szerk.: Molnár Vanda Éva, Debrecen Egyetemi Kiadó, Debrecen, 21-26, 2018. ISBN: 9789633187234

Idegen nyelvű konferencia közlemények (1)

8. **Abriha, D.**: Roofing material determination with hyperspectral data.
In: Student V4 Geoscience Conference and Scientific Meeting GISÁČEK 2017-03-22. Ed.: Jan Tesla, Institute of Geoinformatics, Technical University of Ostrava, Ostrava, 5-10, 2017. ISBN: 9788024840338

További közlemények

Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

9. Szabó, L., **Abriha, D.**, Phinzi, K., Szabó, S.: Urban vegetation classification with high-resolution PlanetScope and SkySat multispectral imagery.
Acta geogr. Debr., Landsc. environ. ser. 15 (1), 66-75, 2021. ISSN: 1789-4921.
DOI: <http://dx.doi.org/10.21120/LE/15/1/9>

Idegen nyelvű tudományos közlemények külföldi folyóiratban (2)

10. Phinzi, K., **Abriha, D.**, Szabó, S.: Classification Efficacy Using K-Fold Cross-Validation and Bootstrapping Resampling Techniques on the Example of Mapping Complex Gully Systems.
Remote Sens. 13 (15), 1-18, 2021. EISSN: 2072-4292.
DOI: <http://dx.doi.org/10.3390/rs13152980>
IF: 5.349
11. Phinzi, K., **Abriha, D.**, Bertalan, L., Holb, I., Szabó, S.: Machine Learning for Gully Feature Extraction Based on a Pan-Sharpned Multispectral Image: Multiclass vs. Binary Approach.
ISPRS Int. Geo-Inf. 9 (4), 1-20, 2020. ISSN: 2220-9964.
DOI: <http://dx.doi.org/10.3390/ijgi9040252>
IF: 2.899





Magyar nyelvű konferencia közlemények (2)

12. Papp, M., Szabó, S., **Abriha, D.**: WorldView-2 és WorldView-3 felvételek értékelése képosztályozáson keresztül.

In: Az elmélet és a gyakorlat találkozása a térinformatikában XIII.: Theory meets practice in GIS. Szerk.: Abriha-Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 255-259, 2022. ISBN: 9789636150396

13. **Abriha, D.**, Szabó, L., Phinzi, K., Szabó, S.: Városi zöldfelületek osztályozása nagy felbontású PlanetScope és SkySat felvételek alapján.

In: Az elmélet és a gyakorlat találkozása a térinformatikában XI. : Theory meets practice in gis. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 13-16, 2020. ISBN: 9789633188866

Magyar nyelvű absztrakt kiadványok (1)

14. **Abriha, D.**, Szabó, S., Varga, G.: Sentinel-5P műholdadatok alapján végzett légköri aeroszol koncentráció vizsgálat Google Earth Engine platformon.

In: Az elmélet és a gyakorlat találkozása a térinformatikában XII.. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi kiadó, Debrecen, 17-23, 2021. ISBN: 9789633189771

Idegen nyelvű absztrakt kiadványok (1)

15. Szabó, S., **Abriha, D.**, Kovács, Z., Sarawut, N., Bertalan, L., Balázs, B.: Pan-sharpening as an effective method to improve classification accuracy of roofing materials.
Geophys. Res. Abstr. 21, 14594, 2019. ISSN: 1029-7006.

A közlő folyóiratok összesített impact faktora: 16,63

A közlő folyóiratok összesített impact faktora (az értekezés alapjául szolgáló közleményekre): 8,382

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudományometriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2023.05.15.



Short thesis for the degree of doctor of philosophy (PhD)

**Application of machine and deep learning
approaches for urban environment analysis using
high-resolution and multidimensional remote
sensing data**

by Dávid Abriha

Supervisor:
Dr. Szilárd Szabó, professor



UNIVERSITY OF DEBRECEN

Doctoral School of Earth Sciences

Debrecen, 2023

Introduction

Due to the growing trend of urbanization in recent decades, the natural environment has been increasingly transformed. The global population growth and the increasing number of megacities (with a population exceeding 10 million) exert greater pressure on the Earth's ecosystems. Such large-scale changes generate numerous economic, social, and environmental challenges, the management of which has become a top priority in many cities. Consequently, sustainability and sustainable urbanization have become one of the most important principles in urban development strategies (Beaverstock et al., 1999; Kadhim et al., 2016; Shen et al., 2012). The focus of my doctoral dissertation is the analysis of the built environment, specifically the analysis of high-resolution and multidimensional remote sensing data of buildings in the city of Debrecen.

Although urban regions are highly complex, remote sensing offers a wide range of opportunities for decision-makers to overcome these challenges. My research has focused on using traditional machine learning methods for image classification and accuracy assessment, as well as applying deep learning approaches to segment buildings.

In my work, I first investigated the separability of different roofing materials under different conditions. I selected the three most common roofing types in a selected study area, and performed classification based on a WorldView-2 satellite image. I then examined the effect of spatial autocorrelation on the accuracy of the models during classification. Prior to the classification of roofing materials, it is advisable to produce a mask that contains only these target objects, thereby reducing the required processing capacity. Additionally, my research aims to examine the accuracy of object detection based on deep convolutional neural networks on relatively inexpensive or easily obtainable images for building extraction.

Aims of the research

During my research. I aimed to find answers to the following questions:

1. What is the effect of the pansharpening process on the classification of roofing materials in WorldView-2 satellite imagery, and how does the separation of individual roof surfaces into sunny and shadowed segments affect thematic accuracy?
2. To what extent does high positive spatial autocorrelation contribute to the overestimation of classification accuracy, and how can the biasing effect of spatial autocorrelation be eliminated to ensure reliability in accuracy estimation through k-fold cross-validation?
3. How effective is the U-Net convolutional neural network for building segmentation using multispectral remote sensing data from different sources with limited training data?

Materials and methods

Image classification and analysis of spatial autocorrelation

I conducted classifications based on roofing materials and accuracy assessments related to spatial autocorrelation using 2 aerial orthophotos (study areas of Debrecen and Vaihingen) and 1 satellite image (WorldView-2; WV2) as part of my dissertation research.

In the case of the Debrecen study area, the necessary reference data for the classification were collected during field surveys, where the types of roofing materials of 350 houses were identified. The classification was based on the three most common roofing materials in the area, namely red tiles, dark tiles (black, grey, brown tiles) and asbestos-cement boards. Although other materials were identified during data collection (metal, tar, etc.), they were not included in the analysis due to their low frequency of occurrence. In the case of the Vaihingen sample area in Germany, Google Earth and Google Street View images were used to identify different types of roofs. As there were no asbestos roofs, but a large number of flat roofs were identified, this became the third category, alongside dark and red tiles.

During the classification, when examining the thematic accuracy of the pansharpening procedure and the separation of roof surfaces based on irradiation, the classifications were carried out using two different approaches. Both traditional multivariate statistical classification methods (linear discriminant analysis [LDFA] and quadratic discriminant analysis [QDFA]) and a robust machine learning algorithm (Random Forest; RF) were employed for the analysis.

Remote sensing images can be strongly affected by spatial autocorrelation (SA), as the values of neighboring pixels are very similar to each other (Wulder and Boots, 1998). Although this is a general characteristic of remote sensing data, it can also question the credibility of the cross-validation results if the selection of the reference data is not appropriate.

To quantify the biasing effect of SA on accuracy, the performance of each model was determined in two ways: (i) the estimated accuracy based on stratified k-fold cross-validation (SCV), and (ii) the true accuracy obtained by validating with independent test data.

In order for cross-validation to be suitable for estimating the true accuracy of the models, a method was developed in Python that selects reference data based on their spatial location: only points outside the specified distance are selected for modeling.

The degree of bias introduced by SA on accuracy was examined for a total of 7 different sampling methods. During validation, accuracy values were determined based on 50 models in all cases, except for the independent testing performed for the true accuracy, which was considered as the control measurement.

Building segmentation based on deep convolutional neural network

I investigated the effectiveness of building segmentation using the U-Net deep learning architecture (Ronneberger et al., 2015) based on 2 satellite images (WV2, WV3) and 1 aerial orthophoto in a study area in Debrecen, Hungary.

Modeling with the U-Net convolutional neural network requires two types of input data: the original raster and the corresponding mask, which in my case is a binary mask that separates buildings from all other objects. Before modeling, the input data was divided into 512 x 512 pixel tiles. For both the orthophoto and WV2 image, all buildings

in an area of approximately 2 km² were vectorized, while for WV3, this area was 0.3 km². The reason for the difference between the two areas is that the evaluation of modeling with fewer training data was also part of the research for WV3.

The manual collection of training data (manual vectorization) is a very time-consuming task. In the course of the research, more than 3000 buildings were vectorized in the three images. Although the U-Net architecture generally works well with relatively few data, augmentation (data expansion) is an essential step to achieve better thematic accuracy. In my research, I applied traditional augmentation methods: the original images were horizontally and vertically flipped, and rotated by 90, 180, and 270 degrees. Thus, in addition to the original image, I obtained 5 additional images.

Each model was trained for up to 50 epochs (the number of training iterations). I implemented an early stopping mechanism that interrupted the training process if the model's performance did not improve for 15 epochs. For each model, I chose an initial learning rate of 0.001. I used the Adam method as the optimization algorithm with a batch size of 20. I used a total of 40 "original" images, i.e., before augmentation, for WV2 and the aerial orthophoto, and 5 and 10 for WV3. Modeling was done using the TensorFlow package (Abadi et al., 2016) in the Python programming language. I used an Nvidia RTX 3090 graphics card with 24 GB VRAM for the modeling.

Scientific results

Thesis statement 1.

In the classification of roofing materials using WorldView-2 satellite imagery, I have found that dividing the categories based on irradiance results in a decrease in thematic accuracy, while improving spatial resolution through pansharpening leads to an increase in accuracy.

For the classification of the WV2 satellite image, I selected 3 categories (asbestos, red tile, dark tile) and classified them using 3

algorithms (LDFA, QDFA, RF). In addition, the analysis was conducted based on the aggregated categories of roofing materials and on the basis of sunny and shadowed areas, resulting in a total of 12 models, including both the original and pansharpened versions of the WV2 imagery (*Figure 1*).

The results showed that the division of categories based on irradiation led to a decrease in accuracy, with OA values being 6-7% lower when sunny and shadowed pixels were separated into different categories. Pansharpening proved to be an effective technique for improving thematic accuracy, with OA values generally increasing by 2-3%, although in some cases (such as the 6 category RF), a difference of up to 6% was observed between the original and pansharpened model versions.

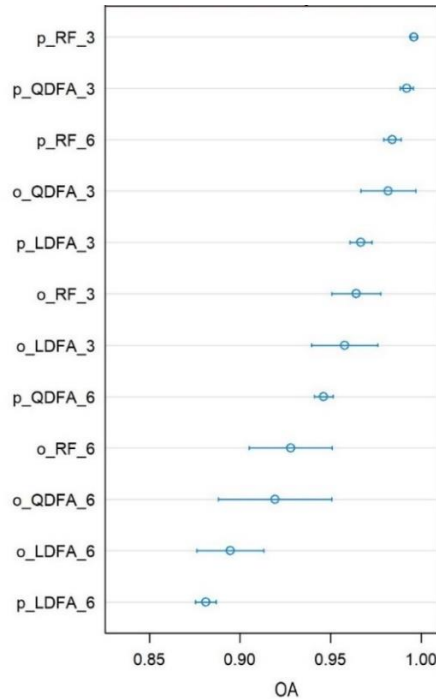


Figure 1. Decreasing rank of Overall Accuracy (OA) of the applied classifiers. – o = original bands; p = pan-sharpened bands; LDFA = Linear Discriminant Function Analysis; QDFA = Quadratic Discriminant Function Analysis; RF = Random Forest; 3 or 6 = number of classes

Thesis statement 2.

I have found that high positive spatial autocorrelation present in the reference data significantly biases the classification accuracy during image classification, which can be effectively eliminated by resampling the data based on their spatial location.

Due to the high spatial autocorrelation, the bias in accuracy estimation was significant when the reference data were randomly split into training and testing sets, as well as when the popular SCV method was applied to estimate the accuracy: in these cases, the discrepancies between the estimated and the true accuracies reached up to 17%, falsely resulting in models with OAs above 99% (*Figure 2*).

However, the method I developed in Python, which selects reference data based on their spatial location, successfully eliminated the distorting effect of SA. The accuracy of the models estimated with SCV and the true accuracies obtained with the independent test data converged well. The lower spatial resolution WV2 image (50 cm) performed 5-9% better than the two orthophotos (9 and 10 cm resolution), mainly because there were fewer pixels representing the roofs in the WV2 image, resulting in reduced intra-class variance within the categories.

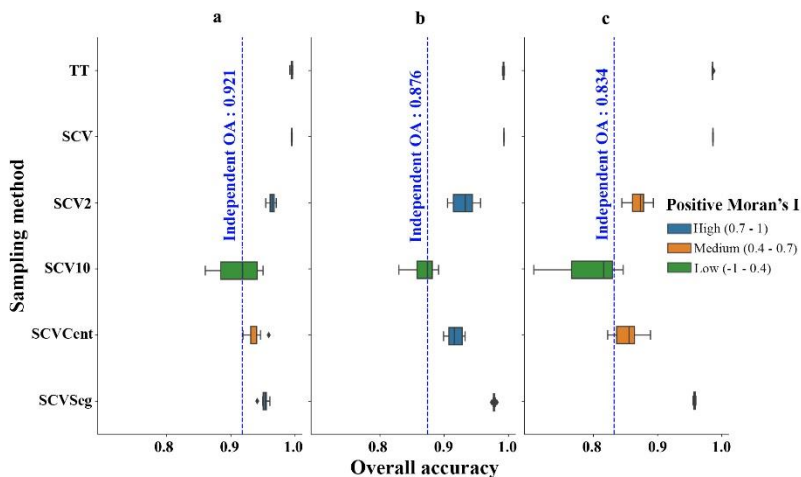


Figure 2. Classification results of the sampling methods based on 50 models with regard to Moran's I (a: WorldView-2; b: orthophoto of Debrecen; c: orthophoto of Vaihingen)

Thesis statement 3.

I have determined that the U-Net deep convolutional architecture can be applied with high accuracy for building segmentation tasks, even with a small number of training data.

When the trainings and predictions were conducted on the same image (2-P40-2; O-P40-O; Figure 3), even with a small amount of training data (192 images, 512 x 512 pixels), the validation accuracy exceeded 94%. However, reducing the amount of training data for WV3 (3-P5-3; 3-P10-3; 24 and 48 images) resulted in overfitting of the learning curves.

If the validation accuracy was high and the model did not overfit, the segmentation was successful (WV2, orthophoto; Figure 4/a. and b.): almost all buildings were identified, with only a few instances of falsely segmented pixels. The buildings incorrectly classified as background (false negatives) were predominantly smaller structures

such as garages and sheds. However, poor segmentation results were obtained for the two models trained on the WV3 image (3-P5-3, 3-P10-3). When utilizing only 24 WV3 training images for modeling (*Figure 4/c.*), a significant portion of the buildings was excluded from the final result. Although some buildings appeared as connected patches during the analysis with 48 training images (*Figure 4/d.*), the overall segmentation performance remained weak and contained numerous errors, with significant misclassification of bare soil surfaces and roads as buildings (false positives).

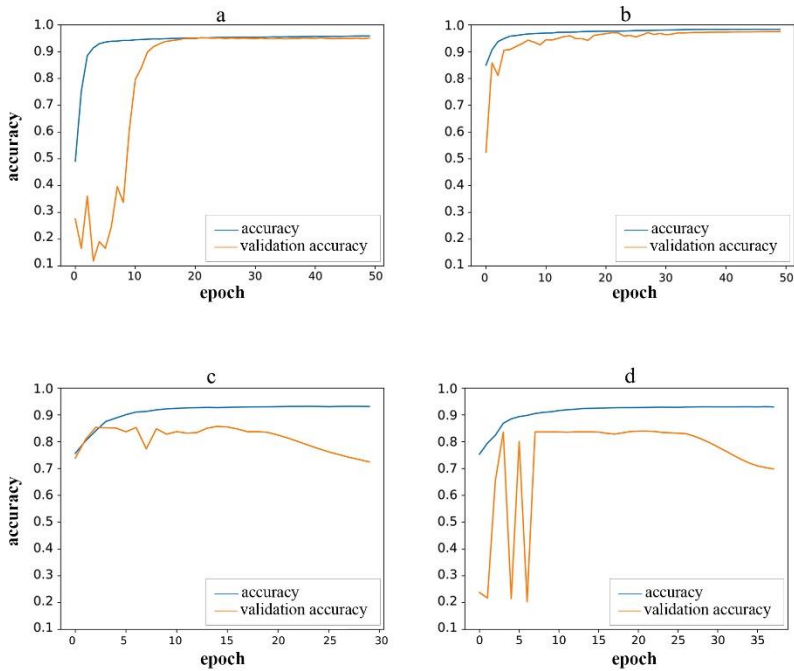


Figure 3. Learning curves in the case where the training dataset and the predicted imagery are the same. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. First and last tags of the labels: 2 = WorldView-2; 3 = WorldView-3; O = Orthophoto.

Second tag = Number of training image used for the prediction (before augmentation)

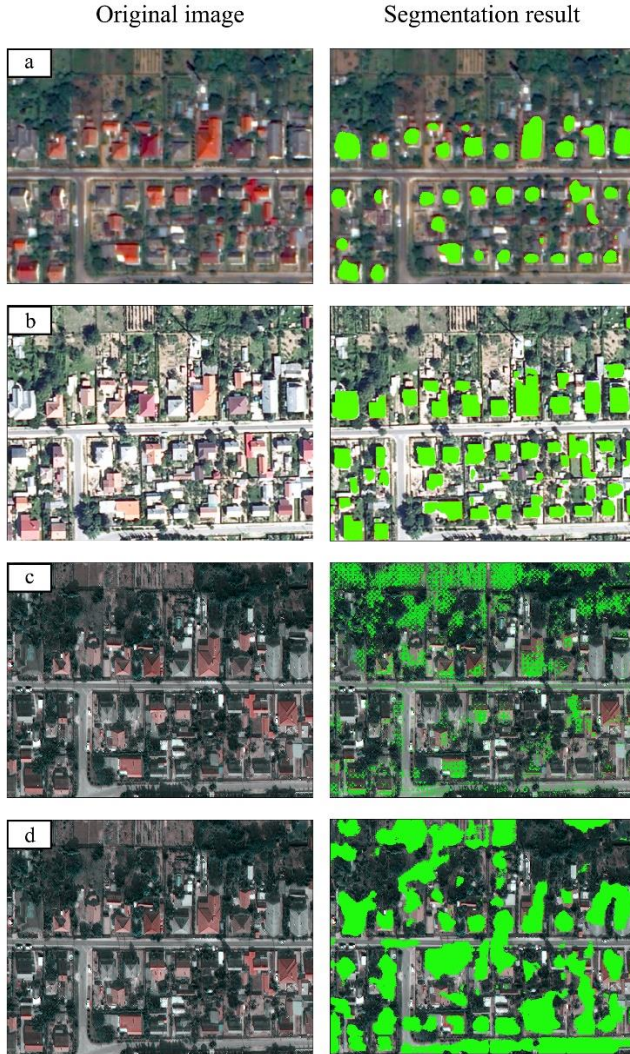


Figure 4. Segmentation results when the training dataset and the predicted image were the same. – a) 2-P40-2; b) O-P40-O; c) 3-P5-3; d) 3-P10-3. First and last tags of the labels: 2 = WorldView-2; 3 = WorldView-3; O = Orthophoto. Second tag = Number of training image used for the prediction (before augmentation)

Thesis statement 4.

I have found that high accuracy can be achieved when fusing remote sensing images from different sources for building segmentation using the U-Net method, provided that at least a small amount of training data from each image is used in the modeling process.

The segmentation of the WV3 image also gave poor results when only the training data of the other two images (WV2, orthophoto) were used for modelling (2O-P80-3): although the number of images used for training increased significantly (384 images), the F1-score was only 0.184. Incorporating 24 images from WV3 (23O-P85-3) significantly improved the accuracy, with an F1-score of 0.661 and visually improved segmentation results. The addition of another 24 images (23O-P90-3; *Figure 5*) further increased the accuracy (F1-score: 0.693; *Table 1*), but to a lesser extent. Although some misclassifications still occurred, their frequency decreased significantly and the model correctly identified the majority of buildings.



Figure 5. Segmentation result based on the combination of all available training data (23O-P90-3). (a: original image; b: segmentation result; first and last tags of the labels: 2 = WorldView-2, 3 = WorldView-3, O = Orthophoto, second tag = Number of training image used for the prediction (before augmentation))

Table 1. Prediction results on the WorldView-3 image based on F1-scores in the case where WorldView-3 images were included for building the models (first and last tags of the labels: 2 = WorldView-2; 3 = WorldView-3; O = Orthophoto. Second tag = Number of training image used for the prediction before augmentation)

Model	Precision	Recall	F1-score
3-P5-3	0.22	0.20	0.213
3-P10-3	0.25	0.72	0.369
23O-P85-3	0.69	0.64	0.661
23O-P90-3	0.68	0.71	0.693

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv:1603.04467.
- Beaverstock, J. V., Smith, R. G., & Taylor, P. J. (1999). A roster of world cities. *Cities*, 16(6), 445–458. [https://doi.org/Doi10.1016/S0264-2751\(99\)00042-6](https://doi.org/Doi10.1016/S0264-2751(99)00042-6), NASA SP-351, NASA, Washington, D.C., pp. 309-317.
- Kadhim, N., Mourshed, M., & Bray, M. (2016). Advances in remote sensing applications for urban sustainability. *Euro-Mediterranean Journal for Environmental Integration*, 1(1). <https://doi.org/10.1007/s41207-016-0007-4>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Shen, L. Y., Peng, Y., Zhang, X. L., & Wu, Y. Z. (2012). An alternative model for evaluating sustainable urbanization. *Cities*, 29(1), 32–39. <https://doi.org/10.1016/j.cities.2011.06.008>
- Wulder, M., & Boots, B. (1998). Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *International Journal of Remote Sensing*, 19(11), 2223–2231. <https://doi.org/Doi10.1080/014311698214983>



Registry number: DEENK/158/2023.PL
Subject: PhD Publication List

Candidate: Dávid Abriha
Doctoral School: Doctoral School of Earth Sciences
MTMT ID: 10063695

List of publications related to the dissertation

Foreign language scientific articles in Hungarian journals (1)

1. **Abriha, D.**, Kovács, Z., Ninsawat, S., Bertalan, L., Balázs, B., Szabó, S.: Identification of roofing materials with Discriminant Function Analysis and Random Forest classifiers on pan-sharpened WorldView-2 imagery - a comparison.
HunGeoBull. 67 (4), 375-392, 2018. ISSN: 2064-5031.
DOI: <http://dx.doi.org/10.15201/hungeobull.67.4.6>

Foreign language scientific articles in international journals (2)

2. **Abriha, D.**, Srivastava, P. K., Szabó, S.: Smaller is better? Unduly nice accuracy assessments in roof detection using remote sensing data with machine learning and k-fold cross-validation.
Heliyon. 9 (3), 1-17, 2023. ISSN: 2405-8440.
DOI: <http://dx.doi.org/10.1016/j.heliyon.2023.e14045>
IF: 3.776 (2021)
3. **Abriha, D.**, Szabó, S.: Strategies in training deep learning models to extract building from multisource images with small training sample sizes.
Int. J. Digit. Earth. 16 (1), 1707-1724, 2023. ISSN: 1753-8947.
DOI: <http://dx.doi.org/10.1080/17538947.2023.2210312>
IF: 4.606 (2021)

Hungarian conference proceedings (4)

4. **Abriha, D.**, Szabó, S., Enyedi, P.: Városi objektum kinyerést célzó deep learning algoritmus alkalmazása nagy felbontású légifelvétel alapján.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XII.. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi kiadó, Debrecen, 9-16, 2021. ISBN: 9789633189771
5. **Abriha, D.**, Szabó, S.: A képosztályozás során fellépő területi autokorreláció vizsgálata Python programozási környezetben.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XI. : theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 17-21, 2020. ISBN: 9789633188866





6. **Abriha, D.**, Banka, F., Szabó, S.: Random Forest osztályozó algoritmus pontosságának vizsgálata tetőfedő anyagok azonosításában multispektrális adatokkal.

In: Az elmélet és a gyakorlat találkozása a térinformatikában X.: Theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 19-24, 2019. ISBN: 9789633180549

7. **Abriha, D.**, Szabó, S.: Hiperspektrális távérzékelés alkalmazása tetőfedő anyagok azonosításában.

In: Az elmélet és a gyakorlat találkozása a térinformatikában IX.. Szerk.: Molnár Vanda Éva, Debrecen Egyetemi Kiadó, Debrecen, 21-26, 2018. ISBN: 9789633187234

Foreign language conference proceedings (1)

8. **Abriha, D.**: Roofing material determination with hyperspectral data.

In: Student V4 Geoscience Conference and Scientific Meeting GISÁCEK 2017-03-22. Ed.: Jan Tesla, Institute of Geoinformatics, Technical University of Ostrava, Ostrava, 5-10, 2017. ISBN: 9788024840338

List of other publications

Foreign language scientific articles in Hungarian journals (1)

9. Szabó, L., **Abriha, D.**, Phinzi, K., Szabó, S.: Urban vegetation classification with high-resolution PlanetScope and SkySat multispectral imagery.

Acta geogr. Debr., Landsc. environ. ser. 15 (1), 66-75, 2021. ISSN: 1789-4921.

DOI: <http://dx.doi.org/10.21120/LE/15/1/9>

Foreign language scientific articles in international journals (2)

10. Phinzi, K., **Abriha, D.**, Szabó, S.: Classification Efficacy Using K-Fold Cross-Validation and Bootstrapping Resampling Techniques on the Example of Mapping Complex Gully Systems. *Remote Sens.* 13 (15), 1-18, 2021. EISSN: 2072-4292.

DOI: <http://dx.doi.org/10.3390/rs13152980>

IF: 5.349

11. Phinzi, K., **Abriha, D.**, Bertalan, L., Holb, I., Szabó, S.: Machine Learning for Gully Feature Extraction Based on a Pan-Sharpned Multispectral Image: Multiclass vs. Binary Approach. *ISPRS Int. Geo-Inf.* 9 (4), 1-20, 2020. ISSN: 2220-9964.

DOI: <http://dx.doi.org/10.3390/ijgi9040252>

IF: 2.899





Hungarian conference proceedings (2)

12. Papp, M., Szabó, S., **Abriha, D.**: WorldView-2 és WorldView-3 felvételek értékelése képosztályozáson keresztül.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XIII.: Theory meets practice in GIS. Szerk.: Abriha-Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 255-259, 2022. ISBN: 9789636150396
13. **Abriha, D.**, Szabó, L., Phinzi, K., Szabó, S.: Városi zöldfelületek osztályozása nagy felbontású PlanetScope és SkySat felvételek alapján.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XI.: Theory meets practice in gis. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 13-16, 2020. ISBN: 9789633188866

Hungarian abstracts (1)

14. **Abriha, D.**, Szabó, S., Varga, G.: Sentinel-5P műholdadatok alapján végzett légköri aeroszol koncentráció vizsgálat Google Earth Engine platformon.
In: Az elmélet és a gyakorlat találkozása a térinformatikában XII.. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi kiadó, Debrecen, 17-23, 2021. ISBN: 9789633189771

Foreign language abstracts (1)

15. Szabó, S., **Abriha, D.**, Kovács, Z., Sarawut, N., Bertalan, L., Balázs, B.: Pan-sharpening as an effective method to improve classification accuracy of roofing materials.
Geophys. Res. Abstr. 21, 14594, 2019. ISSN: 1029-7006.

Total IF of journals (all publications): 16,63

Total IF of journals (publications related to the dissertation): 8,382

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

15 May, 2023

