
Ship Detection in Satellite Images Using Deep Learning

Hajók detektálása műholdképeken mélytanulás segítségével

Sági Benedek

Budapest Műszaki és Gazdaság tudományi Egyetem
benedek.sagi@edu.bme.hu

Czakó Gergő

Budapest Műszaki és Gazdaság tudományi Egyetem
gergo.czako@edu.bme.hu

Abstract

Abstract The automatic analysis of satellite imagery plays a crucial role in maritime traffic monitoring, environmental protection, and security applications. This paper addresses the task defined in the Airbus Ship Detection Challenge, which aims to detect ships in high-resolution satellite images. The proposed approach employs a convolutional neural network-based segmentation model to identify ship locations using binary pixel-level masks. The model is trained in a supervised manner on annotated satellite images, where ship masks are provided in run-length encoded format. During training, separate training and validation sets are used to monitor generalization performance and prevent overfitting. The model is evaluated on training, validation, and test datasets using segmentation-specific metrics and loss functions. Experimental results demonstrate that the deep learning-based segmentation approach is effective for ship detection in complex and noisy visual environments, making it a suitable solution for maritime remote sensing applications.

Magyar kivonat A műholdfelvételek automatikus elemzése kulcsfontosságú szerepet tölt be a tengeri forgalom megfigyelésében, a környezetvédelemben és a biztonsági alkalmazásokban. A dolgozat az Airbus Ship Detection Challenge keretében meghatározott feladatot vizsgálja, amelynek célja hajók detektálása nagy felbontású műholdképeken. A bemutatott módszer egy konvoluciós neurális hálózaton alapuló szegmentációs modellt alkalmaz, amely bináris, pixel-szintű maszkok segítségével határozza meg a hajók elhelyezkedését a képeken. A modellt felügyelt tanulási módszerrel, annotált műholdképeken tanítottuk, ahol a hajókhoz tartozó maszkok futamhossz-kódolt formátumban álltak rendelkezésre. A tanítás során külön tanító és validációs adathalmazt használtunk az általánosítási képesség nyomon követésére és a túltanulás elkerülésére. A modell teljesítményét tanítási, validációs és teszt adathalmazokon értékelte, szegmentációs feladatokra jellemző mérőszámok és veszteségfüggvények alkalmazásával. A kísérleti eredmények azt mutatják, hogy a mélytanulás-alapú szegmentációs megközelítés hatékonyan alkalmazható a hajók detektálására összetett és zajos vizuális környezetben, így alkalmas megoldást jelent a tengeri távérzékelési alkalmazások számára.

1 Bevezető

A műholdas távérzékelés napjainkban kulcsszerepet játszik a Föld felszínének folyamatos megfigyelésében. A nagy felbontású műholdképek lehetővé teszik a tengeri forgalom elemzését, a hajózási útvonalak monitorozását, valamint az illegális tevékenységek - például az engedély nélküli halászat - felderítését. Ezek az alkalmazások megkövetelik a hajók pontos és megbízható automatikus detektálását, amely jelentős kihívást jelent a változatos környezeti feltételek miatt.

A hajók megjelenése a műholdképeken rendkívül heterogén: méretük, alakjuk, orientációjuk és kontrasztjuk jelentősen eltérhet, továbbá a háttér gyakran zajos, hullámzó vízfelszínt vagy part menti struktúrákat tartalmaz. A hagyományos képfeldolgozási módszerek jellemzően nem bizonyulnak elég robusztusnak ilyen komplex vizuális környezetben, ezért az utóbbi években a mélytanulás-alapú megközelítések kerültek előtérbe.

Jelen dolgozat az Airbus Ship Detection Challenge keretében megadott feladatot dolgozza fel, amely hajók pixel-szintű szegmentációját tüzi ki célul műholdképeken. A munka során különböző lehetséges megoldásokat vizsgáltunk meg, beleértve általános célú, előre tanított modelleket és feladatspecifikus, saját fejlesztésű neurális hálózatokat. A végső megoldás egy konvolúciós neurális hálózaton alapuló szegmentációs modell, amelyet kifejezetten az adott adathalmaz sajátosságaihoz igazítottunk.

A hagyományos képfeldolgozási módszerek korlátozott hatékonysságot mutatnak ilyen komplex környezetben, ezért az utóbbi években egyre nagyobb hangsúlyt kapnak a mélytanulás-alapú megközelítések. Jelen dolgozat célja egy konvolúciós neurális hálózaton alapuló modell bemutatása és értékelése az Airbus Ship Detection Challenge adathalmazán, amely alkalmas hajók pontos lokalizálására műholdképeken.

2 Tématerület ismertetése és korábbi megoldások

2.1 Hajódetektálás műholdképeken

A hajók automatikus detektálása műholdképeken a számítógépes látás és a távérzékelés egyik aktívan kutatott területe. A feladat sajátossága, hogy a képek nagy területet fednek le, miközben a detektálandó objektumok gyakran kis méretűek és alacsony kontrasztúak. Emellett a háttér erősen változó lehet, amely tovább nehezíti a megbízható felismerést.

A korai megoldások jellemzően kézzel tervezett jellemzőkön és klasszikus képfeldolgozási eljárásokon alapultak, azonban ezek teljesítménye korlátozott volt komplex vizuális környezetben. A mélytanulás megjelenésével a konvolúciós neurális hálózatok váltak a domináns megközelítéssé, amelyek képesek a releváns vizuális mintázatok automatikus kinyerésére.

2.2 Általános célú szegmentációs modellek

A projekt kezdeti szakaszában megvizsgáltuk az általános célú, előre tanított szegmentációs modellek alkalmazhatóságát. Külön figyelmet fordítottunk a Segment Anything Model (SAM) architektúrára, amely nagyméretű tanítóadatokon betanítva képes különböző objektumok szegmentálására minimális felhasználói beavatkozás mellett.

A SAM dokumentációjának és kapcsolódó publikációinak áttekintése alapján a modell erőssége az általános vizuális reprezentációk kezelése. Ugyanakkor a műholdképek speciális jellemzői - például a hajók kis mérete, a változó felbontás és az alacsony kontraszt - miatt a modell alkalmazása ebben a konkrét feladatban korlátozott hatékonysságot mutathat. Emellett a modell számítási igénye és komplexitása is jelentősnek bizonyult.

Ezeket figyelembe véve végül nem egy általános célú szegmentációs modellt, hanem egy feladatspecifikus, saját tanítású neurális hálózatot választottunk, amely jobban illeszkedik az Airbus Ship Detection Challenge adathalmazához.

2.3 Kaggle közösségi megoldások

A modell tervezése során a Kaggle platformon elérhető nyilvános megoldásokat és kód részleteket is elemeztük. Ezek a közösségi megoldások értékes betekintést nyújtottak az adathalmaz tipikus problémáiba, valamint a gyakran alkalmazott architektúrákba és tanítási stratégiákba.

Számos sikeres megoldás U-Net alapú szegmentációs architektúrát alkalmazott, különböző encoder struktúrákkal és speciális veszteségfüggvényekkel. Gyakori megközelítés volt a Dice-alapú loss függvények használata, valamint az adatok kiegensúlyozásának kezelése, mivel a hajó nélküli képek jelentős túlsúlyban vannak az adathalmazban.

A közösségi megoldások tanulmányozása inspirációként szolgált, azonban a végső modell nem kész megoldások átvételén alapult, hanem a tapasztalatok felhasználásával kialakított, saját implementáció eredménye.

3 Rendszerterv

4 Rendszerterv

A megvalósított rendszer egy végponttól végpontig tanítható, mélytanulás-alapú feldolgozási láncot valósít meg, amely a nyers műholdképekből közvetlenül hajókat jelölő szegmentációs maszkokat állít elő. A rendszer célja, hogy minimalizálja a kézi beavatkozás szükségességét, és automatikusan tanulja meg azokat a vizuális jellemzőket, amelyek a hajók elkülönítéséhez szükségesek.

A bemeneti adatokat RGB műholdképek alkotják, amelyek különböző földrajzi területekről és eltérő környezeti feltételek mellett készültek. A kimenet egy bináris szegmentációs maszk, amely pixel-szinten jelöli a hajókhoz tartozó területeket. Ez a megközelítés lehetővé teszi nemcsak a hajók jelenlétének detektálását, hanem azok pontos lokalizációját is.

A rendszer központi eleme egy konvoluciós neurális hálózat, amely encoder-decoder architektúrát követ. Az encoder feladata a bemeneti képek hierarchikus jellemzőinek kinyerése, amely során a hálózat egyre absztraktabb reprezentációkat tanul meg. A decoder ezekből a reprezentációkból állítja elő az eredeti felbontású szegmentációs maszkot.

Az architektúrában alkalmazott áthidaló kapcsolatok (skip connections) kulcsszerepet játszanak a finom részletek megőrzésében. Ezek a kapcsolatok lehetővé teszik, hogy az encoder korai rétegeiben kinyert alacsony szintű jellemzők közvetlenül eljussanak a decoder megfelelő rétegeibe, ezáltal javítva a szegmentáció pontosságát, különösen kis méretű objektumok esetén.

5 Megvalósítás

5.1 Adatok beszerzése és előkészítése

A feladat során az Airbus Ship Detection Challenge keretében nyilvánosan elérhető Kaggle adathalmazt használtuk. Az adathalmaz nagyfelbontású RGB műholdképeket tartalmaz, amelyekhez hajókat jelölő annotációk tartoznak. Az annotációk futamhossz-kódolással (Run-Length Encoding, RLE) vannak reprezentálva, amely tömör formában írja le a bináris szegmentációs maszkokat.

Az előfeldolgozás első lépéseként az RLE formátumú annotációkat pixel-alapú bináris maszkokká alakítottuk. Ezt követően a bemeneti képeket normalizáltuk, hogy a pixelértékek egységes tartományba essenek, ezáltal stabilabb tanítási folyamatot biztosítva. Az adatok előkészítése során figyelembe vettük azt a sajátosságot is, hogy az adathalmaz jelentős részében nem található hajó, ami osztály-arány eltolódáshoz vezethet.

5.2 Tanítás

A hálózat tanítása felügyelt tanulási módszerrel történt, ahol a bemeneti műholdképekhez tartozó bináris maszkok szolgáltak célértékként. A tanítás során olyan veszteségfüggvényt alkalmaztunk, amely kifejezetten alkalmas szegmentációs feladatokra, és képes kezelní a hajó és nem hajó pixelek közötti egyensúlytalanságot.

Az optimalizálást iteratív, gradiens-alapú módszerrel végeztük, amely során a hálózat paramétereit lépésről lépésre módosítottuk a veszteség minimalizálása érdekében. A tanítási folyamat során az adathalmazt külön tanító és validációs halmazra osztottuk, ami lehetővé tette a modell általánosítási képességének nyomon követését.

A validációs halmazon mért teljesítmény alapján következtettünk a túltanulás megjelenésére, és ennek megfelelően állítottuk be a tanítás hosszát. Ez a megközelítés hozzájárult egy stabilabb és megbízhatóbb modell kialakításához.

5.3 Kiértékelés és tesztelés

A modell teljesítményét a tanítási, validációs és teszt halmazon értékeltük. A kiértékelés során nemcsak a veszteségfüggvény alakulását vizsgáltuk, hanem a szegmentáció minőségét jellemző mérőszámokat is, amelyek pontosabb képet adnak a modell gyakorlati használhatóságáról.

A tanítási és validációs görbék elemzése alapján megállapítható volt, hogy a modell képes volt megtanulni a releváns vizuális mintázatokat, miközben elkerülte a jelentős túltanulást. A teszt halmazon elérte eredmények azt mutatták, hogy a hálózat összetett háttérrel rendelkező képek esetén is megbízhatóan képes a hajók detektálására.

A kiértékelés során külön figyelmet fordítottunk a kis méretű hajók szegmentációjára, mivel ezek jelentik a feladat egyik legnagyobb kihívását. Az eredmények alapján a választott architektúra megfelelő kompromisszumot nyújtott a pontosság és a számítási igény között.

6 Jövőbeli tervezés és összefoglalás

A bemutatott megoldás igazolja, hogy a konvolúciós neurális hálózatokon alapuló szegmentáció hatékony eszközt jelent a hajók automatikus detektálására műholdképeken. A jövőben a modell továbbfejleszthető komplexebb architektúrák alkalmazásával, valamint kiterjedtebb adataugmentációs technikák bevezetésével.

További lehetséges irányt jelent az előtanított hálózatok használata, amelyek gyorsabb konvergenciát és jobb általánosítást eredményezhetnek. Emellett a módszer más távrzékelési feladatokra, például repülőgépek vagy épületek detektálására is kiterjeszthető.

Nyilatkozat mesterséges intelligencia használatáról

A szerzők nyilatkoznak, hogy a dolgozat elkészítése során generatív mesterséges intelligencia alapú eszközöket, közük a ChatGPT-t használtak. Az eszközök elsősorban a szöveg megfogalmazásának finomításához, technikai kérdések tisztázásához, valamint egyes módszertani döntések - például a veszteségfüggvény kiválasztása és a kiértékelési megközelítés - megvitatásához nyújtottak támogatást.

A megoldás koncepcionális tervezését, a neurális hálózat megvalósítását, a kísérletek elvégzését és az eredmények értelmezését a szerzők önállóan végezték. Az AI által generált tartalmak minden esetben ellenőrzésre és szakmai validálásra kerültek.

References

- [1] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- [2] J. Long, E. Shelhamer, T. Darrell. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [3] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [4] F. Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [5] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár. Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

- [6] V. Iglovikov, A. Shvets. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet. *arXiv preprint arXiv:1801.05746*, 2018.
- [7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [8] M. Berman, A. Rannen Triki, M. B. Blaschko. The Lovász-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-over-Union Measure in Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [9] G. Cheng, J. Han. A Survey on Object Detection in Optical Remote Sensing Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.
- [10] S. Li, Z. Zhou, B. Wang. Ship Detection in Optical Remote Sensing Images Based on Deep Learning. *IEEE Geoscience and Remote Sensing Letters*, 2017.
- [11] A. Kirillov et al. Segment Anything. *arXiv preprint arXiv:2304.02643*, 2023.