Ship Detection in Satellite Images Using Deep Learning

Hajók detektálása műholdképeken mélytanulás segítségével

Sági Benedek
Budapest Műszaki és Gazdaságtudományi Egyetem
`benedek.sagi@edu.bme.hu`

Czakó Gergő
Budapest Műszaki és Gazdaságtudományi 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 konvolúciós 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 értékeltük, 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 Introduction

Satellite remote sensing plays a crucial role in the continuous observation of the Earth's surface. High-resolution satellite imagery enables the analysis of maritime traffic, the monitoring of shipping routes, and the detection of illegal activities, such as unauthorized fishing. These applications require accurate and reliable automatic ship detection, which represents a significant challenge due to highly variable environmental conditions.

The appearance of ships in satellite images is highly heterogeneous: their size, shape, orientation, and contrast can vary considerably, while the background often contains noisy water surfaces or coastal structures. Traditional image processing methods generally lack robustness in such complex visual environments, which has led to the increasing adoption of deep learning-based approaches in recent years.

This paper addresses the task defined in the Airbus Ship Detection Challenge, which focuses on pixel-level ship segmentation in satellite images. Several possible approaches were considered during the design phase, including general-purpose pre-trained models and task-specific neural networks. The final solution is based on a convolutional neural network-based segmentation model specifically adapted to the characteristics of the given dataset.

Traditional image processing techniques show limited effectiveness in such complex scenarios; therefore, deep learning-based methods have become increasingly prominent. The goal of this work is to present and analyze a convolutional neural network-based model applied to the Airbus Ship Detection Challenge dataset, capable of accurately localizing ships in satellite images.

2 Related Work and Background

2.1 Ship Detection in Satellite Images

Automatic ship detection in satellite imagery is an actively researched topic at the intersection of computer vision and remote sensing. A key challenge of this task is that images often cover large geographic areas, while the objects of interest are relatively small and exhibit low contrast. Additionally, the background can vary significantly, which further complicates reliable detection.

Early approaches typically relied on handcrafted features and classical image processing techniques; however, their performance was limited in complex visual environments. With the advent of deep learning, convolutional neural networks have become the dominant approach, as they are capable of automatically extracting relevant visual features from data.

2.2 General-Purpose Segmentation Models

During the initial phase of the project, the applicability of general-purpose, pre-trained segmentation models was examined. Special attention was given to the Segment Anything Model (SAM), which is trained on large-scale datasets and is capable of segmenting various objects with minimal user interaction.

Based on the review of the SAM documentation and related publications, the model's main strength lies in handling general visual representations. However, due to the specific characteristics of satellite imagery-such as the small size of ships, varying resolution, and low contrast-the effectiveness of the model for this particular task may be limited. Furthermore, the computational complexity and resource requirements of the model were also considered significant.

Taking these factors into account, a task-specific neural network trained from scratch was selected instead, as it better aligns with the characteristics of the Airbus Ship Detection Challenge dataset.

2.3 Kaggle Community Solutions

During the design of the model, publicly available solutions and code examples on the Kaggle platform were also analyzed. These community solutions provided valuable insights into common challenges of the dataset, as well as frequently used architectures and training strategies.

Many successful approaches employed U-Net-based segmentation architectures combined with different encoder structures and specialized loss functions. Handling class imbalance was a common strategy, as images without ships are present in significantly larger numbers within the dataset.

The analysis of community solutions served as inspiration rather than direct implementation. The final model was developed independently, using the acquired insights to guide architectural and training decisions.

3 System Design

The implemented system represents an end-to-end trainable deep learning-based processing pipeline that produces ship segmentation masks directly from raw satellite images. The primary goal of the system is to minimize manual intervention while automatically learning the visual features required to distinguish ships from the background.

The input consists of RGB satellite images captured under varying geographical and environmental conditions. The output is a binary segmentation mask indicating ship pixels at the pixel level. This approach enables not only the detection of ship presence but also their precise localization.

The core component of the system is a convolutional neural network following an encoder-decoder architecture. The encoder extracts hierarchical feature representations from the input images, learning increasingly abstract features. The decoder reconstructs the segmentation mask from these representations at the original image resolution.

Skip connections play a crucial role in preserving fine-grained details by allowing low-level features from the encoder to be directly propagated to corresponding decoder layers, thereby improving segmentation accuracy, particularly for small objects.

4 Implementation

4.1 Data Acquisition and Preprocessing

The experiments were conducted using the publicly available Airbus Ship Detection Challenge dataset hosted on Kaggle. The dataset consists of high-resolution RGB satellite images along with corresponding ship annotations. The annotations are provided in run-length encoded (RLE) format, which compactly represents binary segmentation masks.

As a first preprocessing step, the RLE annotations were converted into pixel-level binary masks. The input images were then normalized to ensure consistent pixel value ranges, contributing to a more stable training process. Additionally, the significant class imbalance caused by the large number of images without ships was taken into account during data preparation.

4.2 Training

The network was trained using supervised learning, where binary segmentation masks served as target labels for the corresponding input images. A loss function suitable for segmentation tasks was employed, with the ability to handle the imbalance between ship and background pixels.

Optimization was performed using an iterative gradient-based method, where network parameters were updated to minimize the loss function. The dataset was split into training and validation subsets, enabling continuous monitoring of the model's generalization behavior during training.

Based on validation performance, training duration was adjusted to mitigate overfitting. This approach contributed to the development of a more stable and reliable model.

4.3 Evaluation

Model performance was assessed using training and validation subsets. The primary goal of evaluation was to monitor the training process and analyze model behavior, with particular focus on the evolution of the loss function during training.

Analysis of training and validation loss values indicated a gradual reduction in training error, while validation loss remained relatively stable. This suggests that the model was able to learn meaningful visual features without relying solely on memorization of training data.

Qualitative evaluation was performed by visually inspecting the segmentation masks generated for validation images. The results showed that the model was able to identify ship outlines in many cases, although inaccuracies occurred, particularly for small, low-contrast objects or images containing land areas. Overall, the evaluation demonstrates that the chosen architecture is suitable for addressing the core aspects of the task, while still leaving room for further improvements.

5 Future Work and Conclusion

The presented results demonstrate that convolutional neural network-based segmentation is an effective approach for automatic ship detection in satellite imagery. Future work may include the application of more complex architectures and the incorporation of more extensive data augmentation techniques.

Another promising direction is the use of pre-trained networks, which could lead to faster convergence and improved generalization. Additionally, the proposed approach could be extended to other remote sensing tasks, such as the detection of aircraft or buildings.

Declaration on the Use of AI Tools

The authors declare that generative artificial intelligence tools, including ChatGPT, were used during the preparation of this report. These tools were utilized primarily to support language refinement, technical clarification, and the discussion of methodological choices, including the selection of loss functions and evaluation strategies.

The conceptual design of the solution, the implementation of the neural network model, the execution of experiments, and the interpretation of results were carried out independently by the authors. All content generated with the assistance of AI tools was critically reviewed, verified, and adapted to ensure technical correctness and compliance with the requirements of the assignment.

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.