Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/121/2025.19741

A Loss-Modified Densely Connected Atrous Convolutional U-Shape Encoder-Decoder Style Network for Ship Segmentation via Satellite Images

Mingxuan Nie[1,a,*]

[1]*Electrical and Computer Engineering, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, United States*
a. mnie@ucsd.edu
**corresponding author*

Abstract: The segmentation of ships in satellite imagery is vital for maritime surveillance but presents unique challenges such as imbalanced datasets, small-scale object detection, and complex backgrounds. To address these, we propose an advanced U-Net-based architecture enhanced with depthwise separable convolutions for efficient feature extraction, dropout regularization, and batch normalization for improved generalization. Furthermore, a custom combined loss function integrating Dice Loss, Binary Cross-Entropy (BCE), and Focal Loss is introduced to tackle class imbalance and improve segmentation precision. The proposed model achieves superior performance across multiple metrics, including a Dice Coefficient of 0.6325 and a precision of 0.6678, outperforming baseline models such as standard U-Net, Res-U-Net, Dense-U-Net, and MultiRes-U-Net. Qualitative results further demonstrate the model's ability to detect small and complex ship structures, although limitations in noisy environments and misclassification of non-ship objects are observed. Our findings underscore the importance of architectural enhancements and loss optimization in segmentation tasks. Future work could focus on incorporating attention mechanisms, advanced denoising techniques, and multimodal data integration to further improve accuracy in challenging conditions. This study highlights the potential of the enhanced U-Net for maritime applications and broader segmentation tasks in satellite imagery analysis, offering a robust solution for global maritime monitoring.

Keywords: Ship segmentation, Computer Vision, U-Net architecture, Deep Learning

1. Introduction

Satellite monitoring has become an indispensable tool for global maritime surveillance, enabling critical applications such as tracking ships for commercial logistics, national security, and environmental monitoring [1][2][3]. However, ship detection in satellite imagery poses significant challenges due to diverse oceanic conditions, variable lighting, and differences in image resolution. These factors, combined with the small size of many ships relative to the image dimensions, complicate the detection and segmentation tasks.

Traditional object detection methods often fall short in this domain, struggling to balance computational efficiency with detection accuracy. They are particularly inadequate when applied to

DOI: 10.54254/2755-2721/121/2025.19741

large-scale satellite imagery, where both precision and speed are paramount. Recently, deep learning models, especially Convolutional Neural Networks (CNNs), have emerged as powerful solutions for image analysis and segmentation tasks [4][5][6]. Among these, the U-Net model has become a cornerstone for segmentation due to its encoder-decoder architecture, which effectively captures spatial context and local details [7]. However, the standard U-Net is limited in its ability to detect small objects, maintain fine details, and handle imbalanced datasets typical of ship segmentation tasks [8][9][10].

To address these limitations, we propose an advanced U-Net-based architecture specifically tailored for ship segmentation in satellite imagery. Our model leverages several key innovations:

- Depthwise separable convolutions, which enhance feature extraction while significantly reducing computational complexity.
- Batch normalization and dropout regularization, which improve generalization and reduce overfitting.
- A custom combined loss function, incorporating Dice Loss, Binary Cross-Entropy (BCE), and Focal Loss, to effectively address class imbalance and enhance segmentation accuracy.

By integrating these improvements, our model demonstrates superior performance in detecting small and complex ship structures, outperforming baseline architectures such as standard U-Net, Res-U-Net, Dense-U-Net, and MultiRes-U-Net [7][11][12][13]. The proposed approach not only advances the state of the art for ship segmentation but also highlights the broader applicability of these techniques to other challenging segmentation tasks in satellite imagery analysis.

2. Methods

This section details the development of our enhanced U-Net architecture, designed to improve ship segmentation in satellite imagery. Building upon the standard U-Net model, we introduce several modifications to address its limitations in detecting small objects and preserving fine details.

2.1. Baseline U-Net Architecture

The U-Net architecture, introduced by Ronneberger et al. [7], is a fully convolutional network renowned for its efficacy in image segmentation tasks, particularly in biomedical imaging. It comprises two primary components:

- Encoder (Contracting Path): This path captures contextual information through successive convolutional layers, each followed by a rectified linear unit (ReLU) activation and a max-pooling operation [7][14]. The process progressively reduces the spatial dimensions while increasing the feature depth, enabling the network to learn hierarchical features.
- Decoder (Expansive Path): This path reconstructs the spatial dimensions by upsampling the feature maps using transposed convolutions. It incorporates skip connections from the corresponding encoder layers to combine high-resolution features with the upsampled outputs, facilitating precise localization.

DOI: 10.54254/2755-2721/121/2025.19741

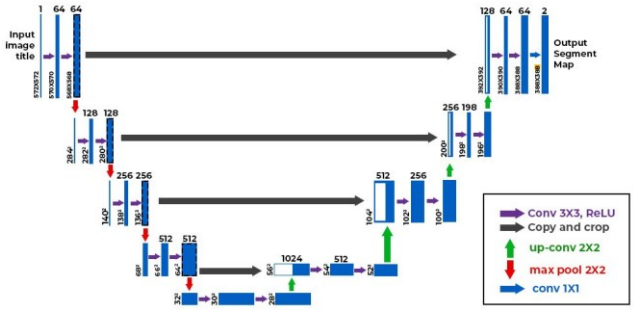

Figure 1: Architecture of the U-Net model

Despite its strengths, the standard U-Net may encounter challenges in accurately segmenting small objects and maintaining fine details, especially in complex backgrounds or when dealing with imbalanced datasets.

2.2. Enhancements in the Proposed Model

To address these challenges, we propose the following enhancements to the U-Net architecture:

- Depthwise Separable Convolutions: Inspired by the MobileNet architecture [15], we replace standard convolutions with depthwise separable convolutions. This factorization divides the convolution into a depthwise convolution, which applies a single filter per input channel, and a pointwise convolution (1×1) that combines these outputs. This approach significantly reduces computational complexity and the number of parameters while retaining the network's capacity to learn rich feature representations.
- Atrous (Dilated) Convolutions: To expand the receptive field without increasing the number of parameters or losing spatial resolution, we integrate atrous convolutions into the network. Atrous convolutions introduce gaps (dilations) between the kernel elements, allowing the network to capture multi-scale contextual information effectively. This technique has been successfully employed in semantic segmentation tasks, as demonstrated by Chen et al. in the DeepLab framework [16].
- Batch Normalization and Dropout: To enhance training stability and mitigate overfitting, we incorporate batch normalization layers after each convolutional operation. Batch normalization standardizes the inputs to a layer for each mini-batch, stabilizing the learning process and enabling the use of higher learning rates [17]. Additionally, we apply dropout regularization, randomly setting a fraction of input units to zero during training, which prevents the network from becoming overly reliant on specific pathways and promotes generalization [18].
- Custom Skip Connections: To improve the preservation of fine-grained spatial details, we modify the skip connections between the encoder and decoder [12][19]. By carefully selecting and integrating features from earlier layers, the network can better retain high resolution information, which is crucial for accurately segmenting small objects such as ships in satellite images.

2.3. Loss Function and Training Strategy

Given the challenges of class imbalance and the need for precise segmentation, we employ a composite loss function that combines Binary Cross-Entropy (BCE), Dice Loss, and Focal Loss:

$$L_{combined} = \alpha \cdot L_{BCE} + \beta \cdot L_{Dice} + \gamma \cdot L_{Focal}$$

where α, β, and γ are weighting factors determined through cross-validation. This combination leverages the strengths of each loss component: BCE focuses on pixel-wise classification, Dice Loss

DOI: 10.54254/2755-2721/121/2025.19741

addresses class imbalance by measuring overlap between predicted and true masks, and Focal Loss further balances the contribution of hard-to-classify pixels [20].

For optimization, we utilize the Adam optimizer [21] with an initial learning rate of 0.0005 and a batch size of 32. To prevent overfitting and ensure robust training, we implement early stopping and learning rate reduction strategies based on validation performance. Evaluation metrics include the Dice coefficient, accuracy, precision, recall, and specificity, providing a comprehensive assessment of segmentation quality.

By integrating these enhancements, our proposed U-Net architecture aims to achieve superior performance in ship segmentation tasks, particularly in detecting small and challenging objects within satellite imagery.

3. Experiments

This section details the experimental setup, including the dataset utilized, implementation specifics, baseline methods for comparison, evaluation metrics, and the results obtained from our enhanced U-Net model.

3.1. Dataset

We employed the Airbus Ship Detection dataset [22], which comprises high-resolution satellite images annotated with ship locations using run-length encoding (RLE) segmentation masks. The dataset includes over 50,000 images, with a significant portion containing no ships. To prepare the data for training, we normalized pixel values to the [0, 1] range and resized all images to a resolution of 128×128 pixels to facilitate efficient training.

3.2. Implementation Details

The model was implemented using Python 3.6 and TensorFlow 2.0. Training was conducted on a single GPU with CUDA support. The learning rate was set to 0.0005, utilizing the Adam optimizer [21]. A batch size of 64 was employed, and the model was trained for 30 epochs. To prevent overfitting and ensure optimal performance, early stopping and model checkpointing were applied to save the best model during training. The total training time was approximately 30 minutes on a single GPU.

3.3. Baseline Methods

For comparative analysis, we evaluated our enhanced U-Net model against several established architectures:

- Standard U-Net: The original U-Net architecture as proposed by Ronneberger et al. [7].
- Res-U-Net: A U-Net variant incorporating residual connections to facilitate deeper network training [23].
- Dense-U-Net: A U-Net modification integrating dense blocks to enhance feature propagation and reuse [13].
- Multi-Res-U-Net: An architecture combining multi-resolution analysis with U-Net to capture features at various scales [12].

All models were trained on the same dataset with identical hyperparameters for fairness.

DOI: 10.54254/2755-2721/121/2025.19741

3.4. Metrics

To assess model performance, we utilized the following metrics:

- Accuracy: The proportion of correctly predicted pixels over the total number of pixels.
- Dice Coefficient: A measure of overlap between the predicted and actual segmentation masks, crucial for evaluating segmentation quality.
- Precision: The proportion of true positive detections out of all positive predictions.
- Recall: The proportion of true positive detections out of all actual positives.
- Specificity: The proportion of true negatives out of all actual negatives.

3.5. Results

The performance of each model is summarized in Table 1. Our enhanced U-Net model demonstrated superior performance across multiple metrics, particularly in the Dice Coefficient and Precision, indicating improved segmentation quality and accurate ship detection.

Table 1: Performance comparison between the standard U-Net and the enhanced U-Net model.

Model	Dice Coefficient	Accuracy	Precision	Recall	Specificity
U-Net [7]	0.6258	0.9989	0.5399	0.7443	0.9992
Res-U-Net [23]	0.2874	0.9963	0.1874	0.6157	0.9975
Dense-U-Net [13]	0.5376	0.9990	0.5736	0.5051	0.9995
Multi-Res-U-Net [12]	0.4813	0.9979	0.3404	0.8214	0.9981
Our Enhanced U-Net	0.6325	0.9992	0.6678	0.6006	0.9996

3.6. Training Metrics and Segmentation Results

Figures 2 through 11 illustrate the training metrics and sample segmentation results for the best baseline model U-Net, and our enhanced U-Net.

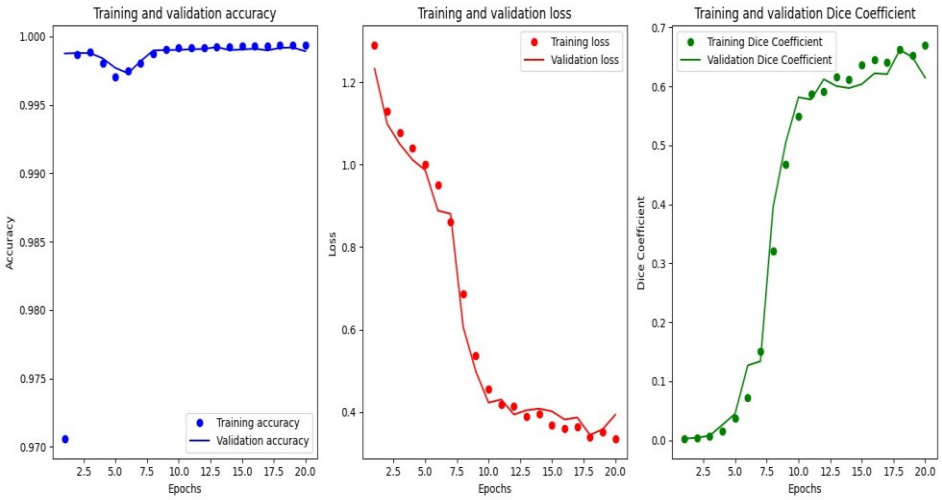

Figure 2: U-Net Training Metrics

DOI: 10.54254/2755-2721/121/2025.19741

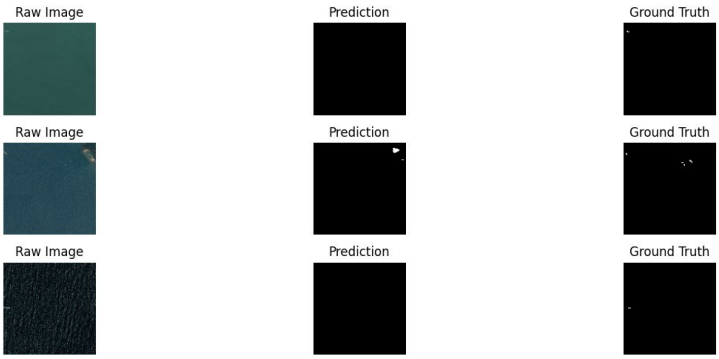

Figure 3: U-Net Segmentation Results

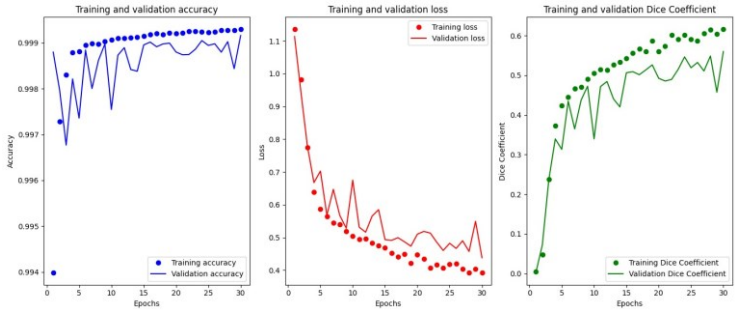

Figure 4: Enhanced U-Net Training Metrics

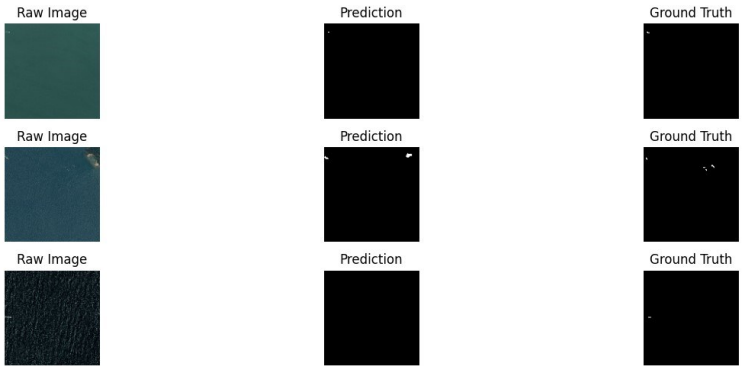

Figure 5: Enhanced U-Net Segmentation Results

3.7. Discussion

The results of our experiments indicate that the enhanced U-Net model achieved significant improvements over the baseline and other comparative models. However, the qualitative analysis of the segmentation outcomes reveals both the strengths and limitations of our model.

As illustrated in Figure 5, the enhanced U-Net successfully identified a small ship in the first example, demonstrating its ability to detect small and challenging objects. This result highlights the effectiveness of depthwise separable convolutions and custom skip connections in preserving fine-grained spatial details.

In the second example, the model misinterpreted a part of an island as a ship. This error suggests that the model occasionally struggles with distinguishing between ships and other objects with similar shapes and textures. Such misclassifications may be attributed to insufficient training examples of

DOI: 10.54254/2755-2721/121/2025.19741

ambiguous cases or the inherent difficulty in differentiating between ships and coastal features in satellite imagery.

The third example shows the model's failure to recognize a ship in a noisy environment, where the sea texture and low contrast obscure the ship. This limitation points to the challenges posed by high noise levels and varying lighting conditions in satellite imagery. While atrous convolutions effectively capture global context, the model's sensitivity to noise remains an area for further improvement.

Future work could focus on addressing these issues by:

- Incorporating more diverse and challenging training samples to improve the model's robustness in distinguishing ships from non-ship objects.
- Utilizing advanced denoising techniques or incorporating additional attention mechanisms to enhance performance in noisy environments.
- Exploring the use of multi-modal data, such as combining optical and radar satellite imagery, to provide complementary information for ship detection.

Despite these limitations, the enhanced U-Net model demonstrated competitive performance, achieving a Dice Coefficient of 0.6325 and precision of 0.6678, as detailed in Table 1. These metrics underline the model's capability to handle complex segmentation tasks, making it a promising approach for maritime monitoring applications.

4. Conclusion

In this paper, we proposed an enhanced U-Net architecture for the task of ship segmentation in satellite imagery. The model builds upon the baseline U-Net by integrating depthwise separable convolutions, atrous convolutions, batch normalization, and dropout regularization, along with a custom loss function combining Binary Cross-Entropy, Dice, and Focal Loss. These enhancements address challenges such as class imbalance, small object detection, and segmentation in noisy environments, yielding notable improvements over the baseline and other comparative models.

Quantitatively, the enhanced U-Net achieved a Dice Coefficient of 0.6325, outperforming the baseline U-Net and demonstrating better segmentation quality. Qualitative results further highlight the model's capability to detect small ships, which are often overlooked by traditional methods. However, some limitations were observed, such as misclassification of non-ship objects and reduced performance in noisy or low-contrast environments. These challenges underscore the complexity of ship segmentation in real-world satellite imagery.

Future research could focus on several directions to address these limitations:

- Increasing the diversity and complexity of the training data to improve robustness.
- Incorporating attention mechanisms or multiscale feature fusion to improve segmentation in noisy or ambiguous regions.
- Exploring multimodal data integration, such as combining optical and radar imagery, to further enhance ship detection accuracy.

Overall, our enhanced U-Net model demonstrates strong potential for maritime monitoring applications, providing a step forward in leveraging deep learning for automatic ship detection. With further advancements, this approach could play a pivotal role in supporting global maritime surveillance and environmental protection efforts.

References

[1] *VG Bondur. Complex satellite monitoring of coastal water areas. In 31st International Symposium on Remote Sensing of Environment. ISRSE, page 7. Citeseer, 2005.*

DOI: 10.54254/2755-2721/121/2025.19741

[2] Shengyang Li, Zhuang Zhou, Manqi Zhao, Jian Yang, Weilong Guo, Yixuan Lv, Longxuan Kou, Han Wang, and Yanfeng Gu. A multitask benchmark dataset for satellite video: Object detection, tracking, and segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61:1–21, 2023.

[3] Aysim Toker, Lukas Kondmann, Mark Weber, Marvin Eisenberger, Andres Camero, Jingliang´ Hu, Ariadna Pregel Hoderlein, C¸aglar S¸enaras, Timothy Davis, Daniel Cremers, et al. Dy-ˇ namicearthnet: Daily multi-spectral satellite dataset for semantic change segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21158–21167, 2022.

[4] Ruirui Li, Wenjie Liu, Lei Yang, Shihao Sun, Wei Hu, Fan Zhang, and Wei Li. Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation. IEEE journal of selected topics in applied earth observations and remote sensing, 11(11):3954–3962, 2018.

[5] Wenqiang Li, Yuk Ming Tang, Ziyang Wang, Kai Ming Yu, and Suet To. Atrous residual interconnected encoder to attention decoder framework for vertebrae segmentation via 3d volumetric ct images. Engineering Applications of Artificial Intelligence, 114:105102, 2022.

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pages 234–241. Springer, 2015.

[8] Sicheng Li. Leveraging deep residual learning with atrous-based unet for enhanced cloud segmentation in satellite imagery. In International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2024), volume 13403, pages 470–475. SPIE, 2024.

[9] Vallikutti Sathananthavathi and G Indumathi. Encoder enhanced atrous (eea) unet architecture for retinal blood vessel segmentation. Cognitive Systems Research, 67:84–95, 2021.

[10] Ziyang Wang and Irina Voiculescu. Quadruple augmented pyramid network for multi-class covid-19 segmentation via ct. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pages 2956–2959. IEEE, 2021.

[11] Sijing Cai, Yunxian Tian, Harvey Lui, Haishan Zeng, Yi Wu, and Guannan Chen. Dense-unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative imaging in medicine and surgery, 10(6):1275, 2020.

[12] Nabil Ibtehaz and M. Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. Neural Networks, 121:74–87, 2020.

[13] Martin Kolaˇr´ık, Radim Burget, Vaclav Uher, Kamil´ Rˇ´ıha, and Malay Kishore Dutta. Optimized high resolution 3d dense-u-net network for brain and spine segmentation. Applied Sciences, 9(3):404, 2019.

[14] AF Agarap. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375, 2018.

[15] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. In arXiv preprint arXiv:1704.04861, 2017.

[16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4):834–848, 2018.

[17] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML), pages 448–456. PMLR, 2015.

[18] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[19] Yu Sun, Fukun Bi, Yangte Gao, Liang Chen, and Suting Feng. A multi-attention unet for semantic segmentation in remote sensing images. Symmetry, 14(5):906, 2022.

[20] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense´ object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2):318– 327, 2020.

[21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[22] Aysim Toker, Lukas Kondmann, Mark Weber, Marvin Eisenberger, Andres Camero, Jingliang´ Hu, Ariadna Pregel Hoderlein, C¸aglar S¸enaras, Timothy Davis, Daniel Cremers, et al. Dy-ˇ namicearthnet: Daily multi-spectral satellite dataset for semantic change segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21158–21167, 2022.

[23] Foivos I. Diakogiannis, Franc¸ois Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing, 162:94–114, 2020.