Landslide Detection and Segmentation Using Remote Sensing Images and Deep Neural Network

Cam Le[1*], Lam Pham[1*], Jasmin Lampert[1], Matthias Schlögl[2,3], Alexander Schindler[1]

Abstract—Knowledge about historic landslide event occurrence is important for supporting disaster risk reduction strategies. Building upon findings from 2022 Landslide4Sense Competition, we propose a deep neural network based system for landslide detection and segmentation from multisource remote sensing image input. We use a U-Net trained with Cross Entropy loss as baseline model. We then improve the U-Net baseline model by leveraging a wide range of deep learning techniques. In particular, we conduct feature engineering by generating new band data from the original bands, which helps to enhance the quality of remote sensing image input. Regarding the network architecture, we replace traditional convolutional layers in the U-Net baseline by a residual-convolutional layer. We also propose an attention layer which leverages the multi-head attention scheme. Additionally, we generate multiple output masks with three different resolutions, which creates an ensemble of three outputs in the inference process to enhance the performance. Finally, we propose a combined loss function which leverages Focal loss and IoU loss to train the network. Our experiments on the development set of the Landslide4Sense challenge achieve an F1 score and an mIoU score of 84.07 and 76.07, respectively. Our best model setup outperforms the challenge baseline and the proposed U-Net baseline, improving the F1 score/mIoU score by 6.8/7.4 and 10.5/8.8, respectively.

Items— Convolutional neural network, landslide, remote sensing image.

I. INTRODUCTION

Natural hazards pose a severe threat to the lives of people around the world. In particular, landslides are a major cause of losses in mountainous areas [1], [2]. Knowledge about historic landslide event occurrence is of core importance in the context of quantitative risk assessment, which in turn supports the design and implementation of effective disaster risk reduction strategies. Several methodological approaches are used for detecting and mapping different types of landslides. In addition to manual visual interpretation, different automated methods that leverage different types of data sets have been developed. Most notably, these methods include the analysis of digital terrain models derived through airborne laserscanning, e.g. by using geographic object-based image analysis [3] or LiDAR altimetry [4], the analysis of aerial photographs [5], or various change detection methods applied to multi-spectral or SAR data [6], [7].

While these methods are tried and tested, the rapid technological development in intersection of remote sens-

ing imagery and image segmentation using increasingly advanced neural network architectures has opened up new possibilities for landslide detection and mapping compared to the conventional methods [8]. The availability of free multi-spectral remote sensing imagery from the satellites, combined with advances in computer vision and machine learning, enables the development of automated landslide detection and segmentation frameworks at comparably low cost.

Recent attempts at developing such systems, which are based on deep neural network architectures such as U-Net, DeepLab, Transformers [9], [10] or on adapted pre-trained models such as variants of ResNet or EfficientNet [11], [12], have presented very promising results. Most of the published systems were based on dedicated datasets collected by the authors [13], [14] or onsynthetic datasets [15]. As a result, these datasets only reflect the landslide events of a certain region, which leads to certain limitations in the developed landslide detection systems.

The Landslide4Sense dataset published by Ghorbanzadeh et al. in 2022 [16], [17] constitutes an interesting and large dataset aimed at landslide detection and segmentation. The data set mainly consists of multi-spectral remote sensing images from Sentinel-2 and (presumably) elevation information as used in the ALOS PALSAR RTC products (i.e., SRTM and NED DEM with geoid correction applied)[1].

Based on the Landslide4Sense dataset, we present a deep neural network based system for landslide detection and segmentation, including the following specific improvements over the benchmark results [16]:

- We first conducted an analysis on how to improve the quality of input remote sensing images by using multiple techniques of data augmentation (random rotation, cutmix) and feature engineering techniques (RGB normalization, feature combination, Gaussian filters, gradient image, Canny Edge detector).
- Second, we improve of the U-Net architecture by proposing a residual-convolutional layer and an attention layer.
- Third, we propose a combination of multi-resolution segmentation heads with multiple loss functions, which also helps to improve model performance.

[1]This information is not really clear from Ghorbanzadeh et al. (2022), who misleadingly state that "DEM and slope data from ALOS PALSAR" [16] were used.

II. Dataset and methodological background

A. Landslide4Sense dataset

The benchmark Landslide4Sense dataset [16] comprises three main subsets: the development set, the evaluation set, and the test set. While the development set was published with the labels, no labels have been provided for both evaluation set and test set as these subsets were used for the competition challenge [18]. Therefore, only the development set of the Landslide4Sense dataset is considered in this paper. This development set comprises 3799 multi-spectral images which were collected from the open source Sentinel-2 [19] and supplemented with information from ALOS PALSAR. Each of multi-spectral image presents 14 bands: multi-spectral data from Sentinel-2 (B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, B12); slope data from ALOS PALSAR (B13); and elevation data (DEM) from ALOS PALSAR (B14). All bands in the dataset have an image size of 128×128. The original spatial resolution of the single bands varies according to the resolution of the source spectral bands of the MSI aboard Sentinel-2: B1, B9 and B10's resolution is 60m, B2 to B4 and B8 were captured at a resolution of 10m per pixel, and B5-B7, B11 and B12 have a resolution of 20m. As a result, each of multi-spectral images is an array of shape $128\times128\times14$. One multi-spectral image of $128\times128\times14$ comes with a label image of 128×128, referred to as the ground truth mask. The ground truth mask presents a binary image in which landslide pixels and non-landslide pixels are marked by one and zero values, respectively. Although approximately 58% of images in the Landslide4Sense development set contains landslide labels, the landslide pixels are minority with only 2.3% of all pixels being labeled as events. Additionally, the ratio of landslide pixels over an image presents a wide range of values from 0.0061% (i.e., only one pixel out of 128×128 pixels in one image) to 47.53% (i.e., nearly a half of pixels in an image). As a result, the dataset presents an imbalance between landslide and non-landslide pixels which causes challenges in the segmentation task.

B. Task definition

Using the development set of the Landslide4Sense [16] dataset as a basis, two tasks of landslide detection and landslide segmentation using deep neural network are proposed in this paper[2]. We evaluate our proposed deep neural networks using random train-test splitting, using 80% for training and 20% as holdout for testing. When the best configuration of the deep neural network is indicated, we evaluate the best network with the 5-fold cross-validation. The final evaluation scores are obtained using the average of scores from 5 folds.

C. Evaluation metrics

Following the guidelines of the Landslide4Sense challenge, we use the F1 score as main evaluation metric [16], [18]. In addition, we report the mean Intersection over Union

[2]The segmentation task was not part of the Landslide4Sense challenge [20].

(mIoU) score, which is a crucial performance metric in the segmentation task [21].

III. Proposed U-Net baseline

The baseline model for for landslide detection and segmentation comprises two main components: Online data augmentation and U-Net based network architecture.

A. Online data augmentation

For the baseline model, we apply two data augmentation methods, rotation and cutmix, to the image input of size $128\times128\times14$. We first randomly rotate each image using an angle of 90, 180, or 270 degrees to generate a new image, referred to as the rotation. Subsequently, random landslide regions from 0 to 2 random landslide images are cut and mixed with the current processing image, referred to as the cutmix [22]. As these data augmentation methods are conducted on each batch of images during training, refer to as step as "online data augmentation".

B. Proposed U-Net based baseline architecture

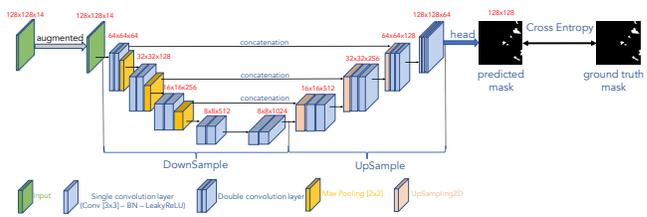The proposed baseline leverages a U-Net architecture (Table I, Fig. 1).

Fig. 1. The U-Net baseline architecture.

The U-Net baseline comprises three main blocks: downsample, upsample, and head. Both downsample and upsample blocks make use of the same double convolution layer. The double convolution layer comprises two single convolution layers, each of which contains one convolutional layer (Conv[3×3]), one Batch Normalization layer (BN) [23], and one Leaky Rectified Linear Unit layer (LeakyReLU) [24]), as shown in Fig. 1. While the downsample block scales down the input images of $128\times128\times14$ to $8\times8\times1024$ by using the Max Pooling layer (MP[2×2]), the upsample block scales up the output of downsample block to $128\times128\times64$ by applying UpSampling 2D. The head block, which uses one dropout layer, one convolutional layer (Conv[1×1]) and applies a Softmax function, helps to transform the output of upsample block to the image of 128×128, referred to as the predicted mask. The predicted mask is compared with the ground truth mask using Cross Entropy as loss function.

We construct the U-Net baseline with the Tensorflow framework. The U-Net baseline is trained for 65 epochs on a an NVIDIA Titan RTX GPU with 24GB RAM. We use Adam optimization [25] for model training.

IV. Improving the U-net baseline system

The improvement of the U-Net baseline focuses on three main aspects of a deep learning model: the loss function, the input quality and the network architecture.

TABLE I

THE U-NET BASELINE ARCHITECTURE

Blocks	Sub-blocks & Layers	Output
	Input	$128\times128\times14$
DownSample	Double convolution layer - MP layer[2×2]	$64\times64\times64$
	Double convolution layer - MP layer[2×2]	$32\times32\times128$
	Double convolution layer - MP layer[2×2]	$16\times16\times256$
	Double convolution layer - MP layer[2×2]	$8\times8\times512$
	Double convolution layer - Single convolution layer	$8\times8\times1024$
UpSample	Upsampling2D layer - Double convolution layer	$16\times16\times512$
	Upsampling2D layer - Double convolution layer	$32\times32\times256$
	Upsampling2D layer - Double convolution layer	$64\times64\times128$
	Upsampling2D layer - Double convolution layer	$128\times128\times64$
Head	Dropout layer(0.2) - Conv layer[1x1] - Softmax	128×128

A. A combined loss function

We tackle the issue of class imbalance between event pixels and non-event pixels by using Focal loss [26]. Additionally, we apply IoU loss [27] to further improve the mIoU score within the segmentation task. As a result, the final loss, referred to as the combined loss, is defined by combining Focal loss and IoU loss with equal weight.

B. Input image quality enhancement

Feature engineering and augmentation are important tuning knobs for improving model performance. We therefore supplement the 14 original bands from the Landslide4Sense development set with 12 additional bands (bands 15 to 26), using methods methods as detailed in Table II.

- Bands 15 to 17 are generated by applying RGB normalization on bands B2, B3 and B4.
- Bands 18 to 21 represent remote sensing indices (NDVI, NDMI, NBR) and a grayscale image.
- Bands 22 and 23 are generated by applying Gaussian and median filters with kernel size of [10×10].
- Bands 24 and 25 are calculated from the image gradient (across length and width dimension).
- Band 26 presents the result of using Canny edge detector.

TABLE II

FEATURE ENGINEERING: ADDITIONAL BANDS

New band data	Formula / Method
Band 15 to Band 17	$(x - x\_min)/(x\_max - x\_min)$
Band 18: NDVI	$(B8 - B4)/(B8 + B4)$
Band 19: NDMI	$(B8 - B11)/(B8 + B11)$
Band 20: NBR	$(B8 - B12)/(B8 + B12)$
Band 21:Gray	$(B2 + B3 + B4)/3$
Band 22 to Band 23	Gausian and Median filters
Band 24 to Band 25	Image gradients across length and width
Band 26	Canny Edge detector

C. U-Net backbone architecture improvement

We propose three main improvements regarding the U-Net baseline architecture. First, we suggest that multiple kernel sizes and a residual based architecture is more effective to capture distinct features of feature maps rather than a conventional convolutional layer. We therefore propose an architecture of a residual-convolutional layer (Res-Conv) which is used to replace the double convolution layer in both the downsample and upsample blocks. Within the proposed residual-convolutional layer (Fig. 2), the input feature map X_1 is first learned by two convolutional layers with

different kernels (e.g. Conv[2×2] and Conv[3×3]) before going through a BN layer, LeakyReLU layer and adding together to generate the feature map X_2. Then, the feature map X_2 goes through a convolutional layer (Conv[3×3]), BN layer, LeakyReLU layer to generate the feature map X_3. Finally, the feature map X_3 is added with the input feature map X_1 to create the final output of the proposed residual-convolutional sub-block.
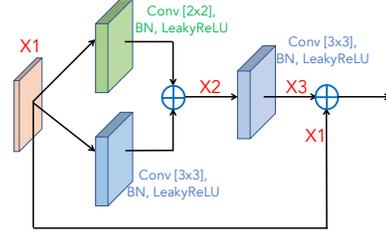

Fig. 2. Residual-convolutional layer.

The second improvement is to apply an attention layer after every convolutional layer in both the downsample and upsample blocks of the proposed U-Net baseline. The attention weights generated by the proposed attention layer effectively support the neural network to focus on landslide regions on the feature maps in the network. We evaluate three types of attention schemes: SE [28] attention, CBAM [29] attention, and multi-head attention [?]. Both SE and CBAM are popular and widely used in literature. Following the line of Le et al. (2023) [30], we propose an additional multi-head attention based layer (Pro-Att) as follows: Given an input feature map X with a size of [$W\times H\times C$] where W, H, and C presents width, height, and channel dimensions, we reduce the size of feature map X across three dimensions using both max and average pooling layers (Fig. 3). The multi-head attention scheme is then applied to each two-dimensional feature maps before multiplying with the original three-dimensional feature map X.
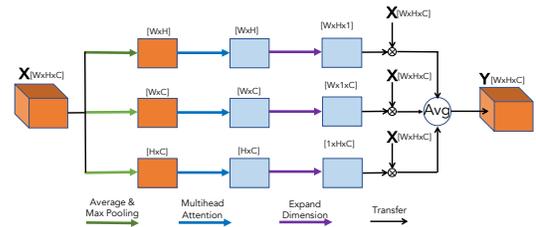

Fig. 3. Attention layer leveraging multi-head attention.

The final improvement is inspired by applying an ensemble of multiple predicted masks with different resolutions to enhance the system performance. In particular, instead of using only one head block to generate one predicted mask of 128×128, we add two more head blocks to generate two other predicted masks: 256×256 and 64×64. As a result, the final predicted result is obtained from an average of three predicted output masks. As we generate three predicted masks, three loss functions are applied for the learning process.

V. Experimental Results

We first evaluate the effect of using the proposed combined loss function using the original images with with 14 bands only. Both Focal loss and IoU loss achieve better performance than the Cross Entropy loss (Table III). The combination of Focal loss and IoU loss yields improvements of 1.22 in the F1-score and 1.13 in the mIoU score, respectively.

TABLE III
Effect of the combined loss function (using 80/20 splitting).

Networks & Loss	F1 score	mIoU
U-Net & Cross Entropy (U-Net baseline)	67.83	60.01
U-Net & Focal Loss	68.28	60.37
U-Net & IoU Loss	68.20	60.23
U-Net & Combined loss	**69.05**	**61.14**

As the proposed combined loss proved to be effective it was set as standard for further evaluation of the newly engineered features. To assess the added value of the new features, the enhanced image input was trained with U-Net baseline using the combined loss. The use of the additional 12 bands leads to further performance improvements by 0.81 in F1-score and 0.62 in mIoU score compared with the U-Net baseline and combined loss (Table IV).

TABLE IV
Effect of feature engineering (U-Net*: U-Net baseline with combined loss function and 80/20 splitting).

Band number	F1 score	mIoU
U-Net* & Original 14 bands	69.05	61.14
U-Net* & Original 14 bands & bands 15 to 17	69.39	61.22
U-Net* & Original 14 bands & bands 15 to 21	69.83	60.97
U-Net* & Original 14 bands & bands 15 to 23	**69.96**	**61.76**
U-Net* & Original 14 bands & bands 15 to 25	69.91	61.64
U-Net* & Original 14 bands & bands 15 to 26	68.54	60.65

We now evaluate the proposed multiple resolution heads, the proposed residual-convolutional layer, and the proposed attention layer. To this end, we use the U-Net baseline, the full 23 band data and the combined loss. All proposed techniques improve the U-Net model performance further (Table VI). While the combination of multiple heads and the proposed attention layer achieves F1/mIoU scores of 71.45/63.05, the combination of multiple heads and the proposed residual-convolutional layer obtains the F1/mIoU scores of 72.07/63.45.

TABLE V
Effect of improving U-Net architecture (U-Net†: U-Net baseline with combined loss function, 23 band data, 80/20 splitting)

Networks	F1 score	mIoU
U-Net†	69.96	61.76
U-Net† & Multiple heads	70.45	62.19
U-Net† & Multiple heads & CBAM Att	70.82	62.53
U-Net† & Multiple heads & SE Att	71.26	62.86
U-Net† & Multiple heads & Pro-Att	71.45	63.05
U-Net† & Multiple heads & Res-Conv	**72.07**	**63.45**

Given the effectiveness of using the combined loss function, the enhanced 23 band data, multi-resolution heads, the proposed Res-Conv layer and attention layers, we eventually configure the best U-Net network architecture (Fig. 4). We evaluate this network with 5-fold cross validation and compare it to the Landslide4Sense challenge baseline as well as the proposed U-Net baseline. The best U-Net network achieves mIoU/F1 scores of 76.07/84.07, thereby outperforming the Landslide4Sense challenge baseline and the proposed U-Net baseline (Table VI). The best U-Net network architecture also performs the lower trainable parameters compared to the other networks.
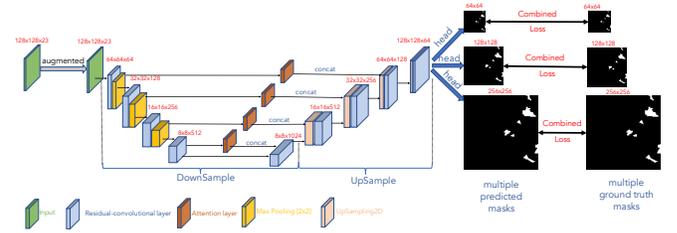

Fig. 4. Proposed optimal U-Net architecture for landslide detection and segmentation using remote sensing imagery.

TABLE VI
Performance comparison between the Landslide4Sense baseline, the proposed U-Net baseline, and the best U-Net network architecture on the Landslide4Sense development set with 5-fold cross validation

Networks	F1 scores	mIoU	Parameters (M)
Landslide4Sense baseline [20]	77.19	68.64	29.8
Proposed U-Net baseline	73.51	67.20	31.0
The best U-Net based network	**84.07**	**76.07**	**24.8**

VI. Conclusion

We have presented a U-Net based deep neural network for landslide detection and segmentation from the remote sensing imagery. We consider and evaluate the effects of improvements of feature engineering, network architecture, and loss functions, and illustrate corresponding improvements in overall network performance. By conducting extensive experiments, we successfully developed an U-Net neural network which achieves an F1-score of 84.07 and an mIoU score of 76.07 on the benchmark Landslide4Sense development set. Our proposed system clearly outperforms the Landslide4Sense baseline by improving the F1-score by 6.88 and the and mIoU score by 7.43, respectively.

ACKNOWLEDGMENTS

REFERENCES

[1] Luciano Picarelli, Suzanne Lacasse, and Ken KS Ho, "The impact of climate change on landslide hazard and risk," *Understanding and Reducing Landslide Disaster Risk: Volume 1 Sendai Landslide Partnerships and Kyoto Landslide Commitment 5th*, pp. 131–141, 2021.

[2] EM-DAT, "The international emergency disasters database," 2023.

[3] Raphael Knevels, Helene Petschko, Philip Leopold, and Alexander Brenning, "Geographic object-based image analysis for automated landslide detection using open source gis software," *ISPRS International Journal of Geo-Information*, vol. 8, no. 12, pp. 551, Dec. 2019.

[4] J McKean and J Roering, "Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry," *Geomorphology*, vol. 57, no. 3-4, pp. 331–351, 2004.

[5] Zhongbin Li, Wenzhong Shi, Ping Lu, Lin Yan, Qunming Wang, and Zelang Miao, "Landslide mapping from aerial photographs using change detection-based markov random field," *Remote Sensing of Environment*, vol. 187, pp. 76–90, Dec. 2016.

[6] Alexander L. Handwerger, Shannan Y. Jones, Pukar Amatya, Hannah R. Kerner, Dalia B. Kirschbaum, and Mong-Han Huang, "Strategies for landslide detection using open-access synthetic aperture radar backscatter change in google earth engine," Oct. 2021.

[7] Simon Plank, André Twele, and Sandro Martinis, "Landslide mapping in vegetated areas using change detection based on optical and polarimetric sar data," *Remote Sensing*, vol. 8, no. 4, pp. 307, Apr. 2016.

[8] GPB Garcia, LP Soares, M Espadoto, and CH Grohmann, "Relict landslide detection using deep-learning architectures for image segmentation in rainforest areas: a new framework," *International Journal of Remote Sensing*, vol. 44, no. 7, pp. 2168–2195, 2023.

[9] Renxiang Huang and Tao Chen, "Landslide recognition from multi-feature remote sensing data based on improved transformers," *Remote Sensing*, vol. 15, no. 13, pp. 3340, 2023.

[10] Han Fu, Bihong Fu, and Pilong Shi, "An improved segmentation method for automatic mapping of cone karst from remote sensing data based on deeplab v3+ model," *Remote Sensing*, vol. 13, no. 3, pp. 441, 2021.

[11] Xinran Liu, Yuexing Peng, Zili Lu, Wei Li, Junchuan Yu, Daqing Ge, and Wei Xiang, "Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[12] Xiaochuan Tang, Mingzhe Liu, Hao Zhong, Yuanzhen Ju, Weile Li, and Qiang Xu, "Mill: channel attention–based deep multiple instance learning for landslide recognition," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 2s, pp. 1–11, 2021.

[13] Lucas Pedrosa Soares, Helen Cristina Dias, Guilherme Pereira Bento Garcia, and Carlos Henrique Grohmann, "Landslide segmentation with deep learning: Evaluating model generalization in rainfall-induced landslides in brazil," *Remote Sensing*, vol. 14, no. 9, pp. 2237, 2022.

[14] Shuai Zhang, Ran Li, Fawu Wang, and Akinori Iio, "Characteristics of landslides triggered by the 2018 hokkaido eastern iburi earthquake, northern japan," *Landslides*, vol. 16, pp. 1691–1708, 2019.

[15] Yongxiu Zhou, Honghui Wang, Ronghao Yang, Guangle Yao, Qiang Xu, and Xiaojuan Zhang, "A novel weakly supervised remote sensing landslide semantic segmentation method: Combining cam and cyclegan algorithms," *Remote Sensing*, vol. 14, no. 15, pp. 3650, 2022.

[16] Omid Ghorbanzadeh, Yonghao Xu, Hengwei Zhao, Junjue Wang, Yanfei Zhong, Dong Zhao, Qi Zang, Shuang Wang, Fahong Zhang, Yilei Shi, Xiao Xiang Zhu, Lin Bai, Weile Li, Weihang Peng, and Pedram Ghamisi, "The outcome of the 2022 landslide4sense competition: Advanced landslide detection from multisource satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9927–9942, 2022.

[17] Omid Ghorbanzadeh, Yonghao Xu, Pedram Ghamisi, Michael Kopp, and David Kreil, "Landslide4sense: Reference benchmark data and deep learning models for landslide detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[18] IARAI, "Landslide4sense," https://www.iarai.ac.at/landslide4sense/challenge/, 2022.

[19] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.

[20] Omid Ghorbanzadeh, Yonghao Xu, Hengwei Zhao, Junjue Wang, Yanfei Zhong, Dong Zhao, Qi Zang, Shuang Wang, Fahong Zhang, Yilei Shi, Xiao Xiang Zhu, Lin Bai, Weile Li, Weihang Peng, and Pedram Ghamisi, "Landslide4sense baseline," https://github.com/iarai/Landslide4Sense-2022, 2022.

[21] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese, "Generalized intersection over union," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[22] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[23] Sergey I. and Christian S., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015, pp. 448–456.

[24] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al., "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, vol. 30, p. 3.

[25] P. K. Diederik and B. Jimmy, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2015.

[26] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[27] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 658–666.

[28] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[29] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[30] Cam Le, Lam Pham, Nghia Nvn, Truong Nguyen, and Le Hong Trang, "A robust and low complexity deep learning model for remote sensing image classification," in *Proceedings of the 8th International Conference on Intelligent Information Technology*, 2023, pp. 177–184.