







Road Extraction From Satellite Images Using Attention-Assisted UNet

Arezou Akhtarmanesh , Dariush Abbasi-Moghadam , Alireza Sharifi , Mohsen Hazrati Yadkouri ,
Aqil Tariq , and Linlin Lu 

I. INTRODUCTION

Abstract—These days, extracting information from remote sensing data has a great impact on various aspects of our lives, such as infrastructure and urban planning, transportation and traffic management, forecasting and tracking natural disasters, searching for mineral resources, monitoring environmental changes, and numerous other fields. One crucial application is extracting accurate road information from aerial images, which has many practical applications ranging from our daily lives to long-term planning for transportation systems to autonomous vehicles. Deep learning models have shown great promise in image-processing tasks, specifically in accurately detecting and extracting roads from aerial images. In this study, various techniques were employed to achieve the desired performance. The model is a UNet assisted with attention blocks in the decoder part and trained with a patched, rotated, and augmented dataset that has been extracted from the DeepGlobe dataset. The preprocessing of the dataset included image and mask patching, rotation, exclusion of background-only images, and excluding images with very little road surface. Both patching and background exclusion in preprocessing as hard attention and attention blocks in the model as soft attention were deployed in order to tackle the inherently biased nature of the dataset. This combination of different techniques empowers the proposed model for superior remote sensing image segmentation performance with an accuracy level of 98.33%. In addition to achieve better performance by the model, another objective is to find the issues that cause the model's performance degradation on some image samples. Therefore, a comprehensive analysis of metrics, with a focus on precision and recall as proper metrics for biased dataset analysis, was conducted to identify potential shortcomings in the model or the dataset, and based on the result, several proposals for future work and further investigations were formulated.

Index Terms—Attention, deep learning, road extraction, satellite images, UNet.

Manuscript received 24 August 2023; revised 8 October 2023 and 4 November 2023; accepted 20 November 2023. Date of publication 28 November 2023; date of current version 8 December 2023. This work was supported by the National Key Research and Development Program of China under Award 2022YFC3800700. (Corresponding authors: Aqil Tariq; Linlin Lu.)

Arezou Akhtarmanesh and Dariush Abbasi-Moghadam are with the Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman 7616913439, Iran (e-mail: arezoo.a70@gmail.com; abbasimoghadam@uk.ac.ir).

Alireza Sharifi is with the Department of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaei Teacher Training University, Tehran 1678815811, Iran (e-mail: a_sharifi@sru.ac.ir).

Mohsen Hazrati Yadkouri is with the Department of Electrical Engineering, Sharif University of Technology, Tehran 1458889694, Iran (e-mail: mohsenhy@gmail.com).

Aqil Tariq is with the Department of Wildlife, Fisheries and Aquaculture, College of Forest Resources, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: at2139@msstate.edu).

Linlin Lu is with the Key Laboratory of Digital Earth Science, Chinese Academy of Sciences Institute of Remote Sensing and Digital Earth, Beijing 100864, China (e-mail: lull@radi.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3336924

ROAD networks are integral components of urban infrastructure, serving as essential transportation arteries that facilitate economic and societal activities [1], [2], [3]. Satellite images are a very important source of remote sensing data, and extracting information such as roads from them has garnered significant attention due to its potential applications in urban planning, autonomous driving, disaster management, traffic analysis, agriculture, environmental monitoring, and smart city construction [4], [5], [6], [7], [8]. However, road extraction from aerial images is a very challenging task due to various reasons, including diverse road types, occlusions, and complex background features. Numerous approaches for extracting roads from remote sensing images have been presented in recent years, and they may be divided into two primary categories: traditional and deep learning-based methods [9]. Traditional approaches primarily rely on image-processing techniques and heuristic strategies to identify road features [10]. These methods involve the recognition of distinct characteristics, patterns, and morphological features, along with the application of handcrafted features and heuristic principles [11], [12]. These traditional approaches are time-consuming, sensitive to variations in road appearance, and often yield limited accuracy in detecting roads from remote sensing images [13]. In recent years, the rapid advancements in satellite imaging technology and the immense expansion of very high resolution (VHR) image data have introduced fresh challenges for road extraction. These challenges arise from the intricate structure of roads and the diverse distribution of backgrounds. Given these factors, traditional methods have shown a decline in performance as they struggle to achieve effective generalization and are constrained in expressing essential features. Considering the large volume of images and the diverse patterns within them, machine learning-based methods are more suitable for their automatic nature and reduced reliance on subjective feature selection [13]. Among the different machine learning methods, deep learning techniques are preferred for road extraction tasks because of their ability to automatically extract image features from vast datasets and eliminate the need for manual feature selection [14], [15], [16]. Deep learning techniques have recently emerged as very powerful tools for a variety of tasks, including road extraction tasks, leveraging the abundance of labeled data, and the capacity of deep neural networks to learn discriminative features automatically [9].

Convolutional neural networks (CNNs), as a branch of deep learning, are the most suitable method with superior

performance for road extraction and image-processing tasks. Generally, CNNs are well-suited for array-like data, such as images and time series [17], [18], [19], [20], [21], [22], [23]. For example, Zhong et al. [24] introduced a CNN model that effectively combines low-level fine-grained features with high-level semantic features, enabling the extraction of road and building targets in satellite imagery. Patch-based convolutional network, as an evolution of CNNs in the context of road extraction, was implemented in 2010 [25]. Alshehhi et al. [26] proposed a patch-based CNN model capable of simultaneously extracting road and building components from remote sensing imagery. Subsequently, the fully convolutional network (FCN) [27] model-based road extraction method emerged. Varia et al. [28] utilized the FCN-32, a deep learning technique, for extracting road segments from extremely high-resolution unmanned aerial vehicle (UAV) imagery. Building upon FCN, Kestur et al. [29] presented the U-shaped FCN, tailored for road extraction from UAV images. In another approach, Panboonyuen et al. [30] utilized landscape metrics and the exponential linear unit functions to extract road objects from remote sensing imagery. Hong et al. [31] applied a block based on richer convolutional features to segment roads from high-resolution remote sensing imagery. Cheng et al. [32] proposed a cascaded end-to-end CNN model for extracting road centerlines from remote sensing imagery. The strength of CNN-based models lies in their ability to automatically discern road characteristics, leveraging strong generalization, arbitrary function fitting, and high stability. CNN-based models consistently demonstrated superior performance compared to the traditional methods. However, it is important to note that they heavily rely on an ample quantity of images and the availability of remote sensing images is typically limited.

In recent years, to segment roads with limited remote sensing images, architectures inspired by the encoder–decoder model, such as UNet, have gained prominence in road segmentation tasks. Building upon the FCN [27] and encoder–decoder model ideas, the UNet model was developed specifically for biomedical image segmentation at first [33]. The UNet architecture takes inspiration from FCNs and employs an encoder-decoder structure that enables the capture of both local and global contexts, leading to effective segmentation [34]. Various extensions and adaptations of UNet have been developed to address specific challenges in road extraction, such as SegNet [35] and Res-UNet [36], which is an enhancement of the UNet architecture achieved by integrating a residual module [37]. Another innovation by Zhou et al. [38] introduced the D-LinkNet34 model, which, building upon the UNet, expanded the receptive field while preserving resolution by simultaneously employing a dilated convolution module [39].

Numerous other networks have also been devised and realized employing the encoder–decoder architecture. In 2020, a road segmentation network named global context-based automatic road segmentation via dilated convolutional neural network employed dilated convolutions in an encoder–decoder architecture. It adopted a structure akin to UNet, integrating three residual dilated blocks within the encoder to expand the receptive field

[40]. The C-UNet, which was proposed by Hou et al. [41], incorporates a complement module for better feature representation in road extraction from remote sensing images, despite increased complexity. Recently, Yang et al. [42] presented SDUNet, a model that suggests spatial enhancement and densely connected blocks within the UNet framework to improve the precision of road extraction.

Attention-based models have also been successful in improving road extraction accuracy. The attention module focuses on important image parts for accurate results, leveraging patterns and spatial relevance, and modifies CNN architectures by generating score matrices from intermediate feature maps, enhancing relevant features and suppressing noise, leading to improved classification across datasets [43], [44], [45]. A notable case among attention-based models is the global and local attention model based on U-Net and DenseNet. This model, based on the UNet architecture, incorporates region-based and global attention methods in its design [46]. Ren et al. [47] DA-CapsUNet combine capsule representations and attention mechanisms for robust feature fusion in road region extraction. Dual-attention network (DA-RoadNet), proposed by Wan et al., utilizes a shallow encoder–decoder network with densely connected blocks to safeguard road structure information. It integrates a novel attention mechanism and a hybrid loss function to address class imbalance effectively [48]. Li et al. introduced Cascaded Attention DenseUNet (CADUNet), a model that incorporates global and core attention modules within DenseUNet architecture. This integration enhances road network connectivity and ensures the preservation of the integrity of shaded road areas [49]. Li et al. [50] cascaded attention-enhanced architecture includes spatial and channel attention for adaptive boundary refinement in road extraction. Shao et al. presented a road extraction convolutional neural network with an embedded attention mechanism (RENA), combining channel and spatial attention within the U-Net framework. The network employs residual densely connected blocks for improved feature reuse and information flow, complemented by a residual dilated convolution module for multiscale road network extraction [51].

Despite significant advancements facilitated by deep learning methods and innovative techniques, achieving highly accurate models for road extraction from remote sensing images remains a challenge. The inherent bias in road extraction datasets necessitates solutions that address this issue effectively. Attention mechanisms and the UNet architecture prove to be promising solutions by focusing on crucial parts of the image and capturing both local and global contexts, respectively. To further enhance performance, we implemented a UNet model equipped with an attention block as soft attention and used patching and background exclusion in preprocessing as hard attention to achieve better performance. Additionally, we conduct a comprehensive analysis, focusing on precision and recall as suitable metrics for skewed datasets to gain deeper insights into the model's performance.

In the sections that follow, we will go into more detail about our methodology, including the network architecture and strategies we used, which are covered in Section II. The results and

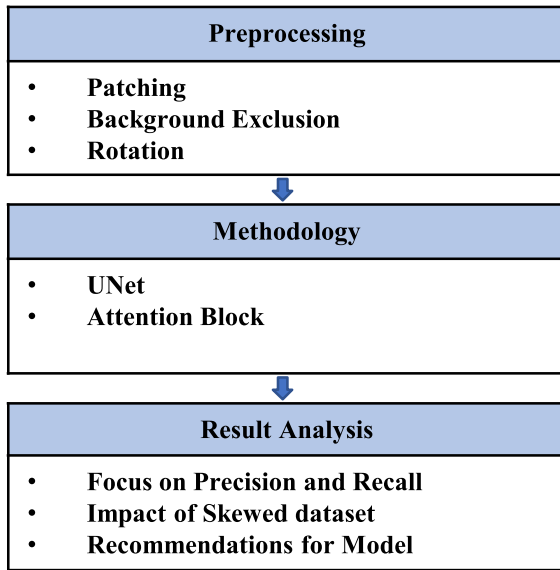


Fig. 1. Flowchart of the proposed method.

configurations utilized for result analysis are then thoroughly discussed in Section III. Conclusions and recommendations for additional research are provided in Section IV.

II. MATERIALS AND METHODS

The overall procedure of this work can be divided into three parts: preprocessing of the dataset, model implementation, and analysis of the result. These three steps and the considerations in each step have been shown as a flowchart in Fig. 1. By patching, background exclusion, and rotation in preprocessing and deploying UNet and attention blocks in the model, we have tried to overcome the biased dataset issue and achieve better performance. Finally, in the result analysis step, we have focused on precision and recall to have a better evaluation of the performance and find out what the shortcomings are that should be considered for future works.

The dataset for road extraction is inherently very skewed, as only a small portion of the earth's surface is covered by roads. In the road extraction task, it is very important to overcome this issue. We have considered it in all steps of the work, from data preparation to result analysis. Attention as a technique to tackle the skewed dataset issue and improve performance has been deployed in both hard and soft attention types. In data reparation, after patching the images, the background-only images were excluded, which is a kind of hard attention and increases the road portion in the training set. The attention blocks deployed in the UNet model perform as soft attention, which learns to pay attention to the areas covered by roads. The metrics, including precision, recall, Intersection over Union (IOU), and F-Score, were measured on the test set, and then deeper analysis of the result was done with a focus on precision and recall as these are the proper metrics for biased dataset analysis. By these techniques and considerations, the model achieved superior performance compared to most of the models implemented based on UNet.

A. Dataset

In this study, we utilized the DeepGlobe dataset [52]. This dataset contains 6226 image-mask pairs with a size of 1024×1024 and a resolution of 0.5 m as a training set, covering areas of Thailand, India, and Indonesia. The roads encompass cement, asphalt, and mountain roads. The image format includes RGB color images, which are characterized by multispectral or hyperspectral bands, encompassing essential wavelengths such as red, green, and blue. These spectral bands enhance feature discrimination and analytical capabilities, crucial for tasks such as road extraction. The DeepGlobe dataset comes with clear limitations and potential biases. One major concern is class imbalance, where the presence of significantly more nonroad pixels compared to road pixels could divert the model's attention away from important road features during training. Another potential issue is a geographic bias within the dataset, possibly favoring regions with more abundant or easily accessible data. This bias might limit the model's ability to effectively generalize across various geographical landscapes. Furthermore, variations in image resolutions pose a challenge, especially in accurately detecting smaller or less pronounced road features. In addition, inconsistencies in annotation quality, including discrepancies or inaccuracies in road markings, can adversely affect the model's training and overall performance. These identified limitations underscore the critical need for careful consideration and the implementation of suitable mitigation strategies when utilizing this dataset for road extraction tasks. The website for the DeepGlobe dataset can be found at: <https://deepglobe.org/challenge.html>.

The validation and test images of the DeepGlobe dataset are not masked, so we kept 1726 image-mask pairs of the dataset as a test set and used the remaining 4500 images as training and validation images after some preparation and augmentations described in the following.

In order to prepare the training set, each training image was patched into nine overlapping 512×512 images. Patching the images caused some of the new images to be completely laid in the background area, with some of them including a negligible portion covered by road. These two categories contributed almost 25% of the total patched images. Considering that the learning process of deep neural networks primarily involves identifying similarities and differences between the classes, these images, which are mainly composed of one class, did not contribute positively to the learning process and might even have led to poor learning of the model. Excluding these images would cause better learning of the model in addition to mitigating the bias issue in the dataset. Omitting these 10 000 images from the patched images, we finally applied rotations of 0° , 90° , 180° , and 270° to the remaining images, resulting in a quadrupling of the dataset. Subsequently, we split these augmented images into a training set (105 000 images) and a validation set (15 000 images) to train the model.

B. Methodology

We implemented a UNet architecture assisted with attention blocks in the decoder part to extract roads from the dataset images by classifying pixels into two categories: road and

background. This model has 2.07 million trainable parameters. The input image size is 512×512 which halves through a max-pooling layer after each convolutional block. Simultaneously, the depth of the feature map matches the number of filters utilized in that specific convolutional block. Within the encoder part, every convolutional block comprises two convolutional layers. The filter count of each convolutional block compared to the preceding block is doubled. Starting with 16 filters at the first block and doubling at each block, at the model's bottom layer, the feature map reaches a depth of 512 with the size of 16×16 . After each max-pooling layer, a dropout layer is considered which is incorporated at a rate of 0.1 in the initial layer, with a gradual increase to 0.5 at the deepest, bottleneck layer. This strategic approach aims to counteroverfitting, enhance generalization by introducing controlled noise during training, and prevent an overly tailored fit to the training data [53].

Due to the substantial bias in our dataset, where the background area vastly outnumbered the road area, we incorporated the Attention block to enhance the model's focus on regions containing roads. Spatial attention enhances accuracy by capturing intricate input–output relationships, promoting scalability across input sizes, and also improving interpretability by visualizing the attended input parts and their impact on output [54]. An attention block is a component that filters out irrelevant features for the current task. In the expansive path, each layer incorporates an attention gate, requiring features from the encoder path to pass through it before concatenating with up-sampled features [20], [55]. The attention block receives two inputs: one from the deeper layer referred to as the “gating” signal and one from the encoder part referred to as the “skip” signal. Due to their different layer origins, their dimensions differ. Both signals pass through a convolutional layer, and the gating signal is up-sampled. The resulting equal-sized outputs are then added together, incorporating deep feature map information from the gating signal and spatial information from the skip signal. This combinational feature map is passed through a ReLU activation layer followed by a single filter convolutional layer and then a sigmoid activation layer, resulting in a 2-D array of numbers that can be considered as attention weights. This array multiplied by the skip signal is used to weight the skip signal, signifying the target areas that contain roads and attenuating other regions. Attention block applies a function that weights feature maps based on class importance, allowing the network to prioritize specific classes and focus on particular objects within an image. The weighted skip signal is then concatenated with an up-sampled gating signal to form the input for the subsequent convolutional layer. Note that the number of filters to be convolved with skip and gating signals is equal to the depth of the feature map of the skip signal, which is different for each attention block.

The implemented model is shown in Fig. 2. Note that the arrows representing convolution, down-sampling, up-sampling, and skip connections are shown with different colors. The up-sampling process involves a twofold approach. Initially, there is an up-convolution, which enlarges the feature map by doubling its width and height while reducing the channel count by half. Following this, the enlarged feature map is merged with the corresponding feature map from the contracting path (encoder)

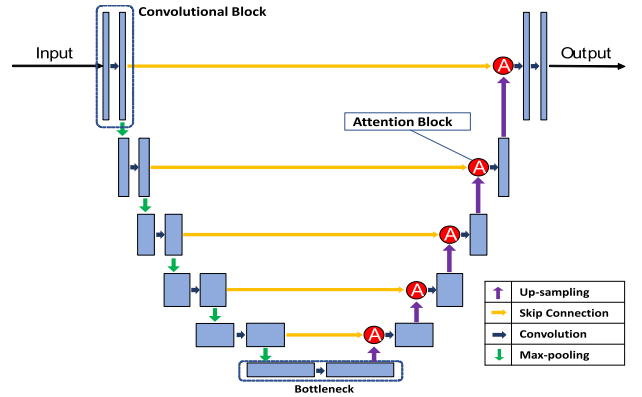


Fig. 2. Proposed architecture of UNet with attention block.

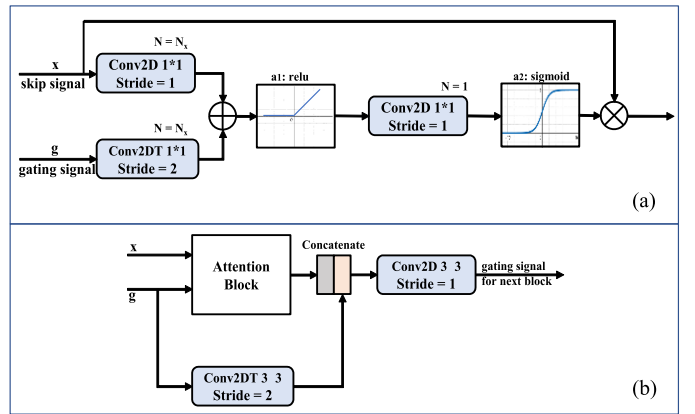


Fig. 3. (a) Detailed schematic of implemented attention. (b) Attention block's position in the model.

through concatenation. This maintains the spatial information and context. Consequently, the up-sampling operation not only reinstates the resolution but also integrates the acquired low-level and high-level features within the network [33]. Skip connections address the vanishing gradient issue by preserving gradient magnitude, encouraging flat minimizers, and stopping the transition to chaotic behavior [54]. Moreover, they enhance model accuracy by amalgamating low-level and high-level features, enabling the capture of intricate patterns and improving generalization to new data. In addition, skip-connections combat overfitting by introducing noise to feature maps, thereby promoting more resilient model architecture [56]. The attention block, which is the main block of this model, is shown in Fig. 3. In the attention block, the additional layer adds the gating signal and skip signal, and the multiplication layer multiplies the attention weights by the skip signal, resulting in the weighted feature map.

The additive attention block is described as follows:

$$q_{\text{att}}^1 = \psi^T \left(\sigma_1 \left(W_x^T x_i^l + W_g^T g_i + b_g \right) \right) + b_\psi \quad (1)$$

$$\alpha_i^l = \sigma_2 \left(q_{\text{att}}^1 \left(x_i^l, g_i; \Theta_{\text{att}} \right) \right) \quad (2)$$

where x^l is the feature from the contracting path and g is the gating signal. Also, the term $\sigma_2(x_{i,c}) = \frac{1}{1 + \exp(-x_{i,c})}$ corresponds to the sigmoid activation function.

Attention block is characterized by a set of parameters Θ_{att} containing linear transformations $W_x \in R^{F_l \times F_{\text{int}}}$, $W_g \in R^{F_g \times F_{\text{int}}}$, $\psi \in R^{F_{\text{int}} \times 1}$, and bias terms $b_\psi \in R$, $b_g \in R^{F_{\text{int}}}$. The input tensors' channelwise $1 \times 1 \times 1$ convolutions are used to compute the linear transformations. The concatenated features x^l and g are linearly transferred to an $R^{F_{\text{int}}}$ dimensional intermediate space [56], and this is known as vector concatenation-based attention in other settings [45].

It is worth noting that for training the model, we employed the Adam optimizer in conjunction with the binary cross-entropy (BCE) loss function, which is suitable for binary classification tasks and is frequently employed to quantify the difference between two probability distributions. A one-hot encoded vector that has the true class's value set to 1 and the other class's value set to 0 is typically used to represent the true class in binary classification [57], [58]

$$\text{BC E}_{\text{Loss}} = -\frac{1}{N} \sum_{i=0}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (3)$$

where p is the anticipated probability of the positive class and y is the true class label (either 0 or 1). When the predicted probability p equals the actual class label y , the loss function is minimized. Numerous advantageous characteristics of the BCE loss include its simplicity, differentiability, and ability to interpret the results of the model in probabilistic terms. In addition, compared to other loss functions, it offers a smooth optimization surface and is less susceptible to outliers [58].

Also, in the output layer, the sigmoid function was used as activation. Apart from the final layer, in each Attention block, we also used a sigmoid activation layer to create the Attention array [59]

$$\text{Sigmoid activation function} = \frac{1}{1 + e^{-x}}. \quad (4)$$

For the remaining layers, the ReLU activation function was applied. The reason behind using the sigmoid activation function in the mentioned layers is its effectiveness in improving class separation. In addition, the utilization of ReLU activation in other layers aims to expedite parameter convergence, thus accelerating the learning process.

The ReLU nonlinear function's primary benefit is that it has a stable derivative for all input values greater than zero. Network learning is accelerated by this fixed derivative. Levels are primarily utilized to extract important features that will be applied to subsequent levels to carry out the classification process [52]

$$\text{ReLU activation function} = \max(x, 0). \quad (5)$$

C. Accuracy Assessment

Despite the training step, where we divided each image into nine overlapping smaller patches as an augmentation technique and excluded the background-only images, in the evaluation step, we considered full coverage of the image for analyzing the model's performance on the test set. The metrics precision, recall, F1-score, and IOU were used to evaluate the trained model on the test set. The threshold value of 0.5 is used to classify each

pixel of the output into the road or background class. Considering that each pixel is labeled as road or background in the mask and the same happens in the prediction, there are four types of pixels. The road pixels of the mask, which are predicted as the road by the model as well, are called true positive (TP). The road pixels of the mask, which are predicted as background by the model, are called false negative (FN). The same happens for the background pixels of the mask. The background pixels predicted as negative are called true negative (TN), and the remaining background pixels predicted as positive are false positive (FP). By counting the number of TP, TN, FP, and FN, the metrics are calculated.

Precision measures the proportion of correctly predicted road pixels out of all the pixels predicted as road. Low precision means that a significant portion of the pixels predicted as road pixels are not really road pixels. In other words, low precision means that the background pixels predicted as road are considerable compared to the real road pixels predicted as road.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}. \quad (6)$$

Recall measures the proportion of correctly predicted road pixels out of all actual road pixels. By actual road pixels, we mean the road pixels of the mask. Recall is also known as sensitivity or TP rate. Low recall means that a considerable portion of actual road pixels are predicted as background.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (7)$$

There is usually a tradeoff between precision and recall. In order to achieve high precision, FPs should be minimized. Minimizing FP means the model is cautious in labeling the pixels as roads. Being very cautious, the model may even not label the real road pixels as road, which would result in an FN increment and hence low recall. The metric F-Score is usually used in order to consider the precision-recall tradeoff.

$$\text{F-Score} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})}. \quad (8)$$

Another important metric evaluated on the test set is IOU, which measures the intersection of actual and predicted road pixels over all the actual and predicted road pixels.

$$\text{IOU} = \frac{\text{TP}}{(\text{TP} + \text{FN} + \text{FP})}. \quad (9)$$

All these metrics range from zero to one, with higher values indicating better performance. Precision signifies the impact of FPs, and recall signifies the consequences of FNs. The F-Score and IOU include both FPs and FNs. In order to compare performance considering a single metric, the F-Score and IOU are proper metrics, but precision and recall usually should be considered simultaneously, and considering both of them would be useful for a more detailed analysis of performance.

III. RESULTS AND DISCUSSION

The model was trained on NVIDIA A100 GPU with 40 GB GRAM on the Google Colab platform. In this work, we followed two goals: 1) improve the performance of road extraction by the

TABLE I
RESULT COMPARISON OF ROAD EXTRACTION METHODS PERFORMED ON
DEEPGLOBE DATASET

Method	Precision	Recall	F1	IOU
UNet [33]	75.13	55.53	62	46.3
SegNet [35]	54.45	76.38	60	45.45
Res-UNet [36]	84.97	64.13	71.95	57.36
D-LinkNet [38]	77.74	78.65	77.04	63.98
DA-RoadNet [48]	73.65	69.55	71.54	55.7
CADUNet [49]	74.89	78.66	76.28	62.08
SDUNet [42]	78.4	74.2	79.4	66.8
Cascaded Residual Att. [50]	79.12	80.39	78.75	66.38
RENA [51]	78.4	77	76.4	63.1
Proposed Method	79.95	81.15	80.54	67.42

The bold font indicates best result obtained for analysis.

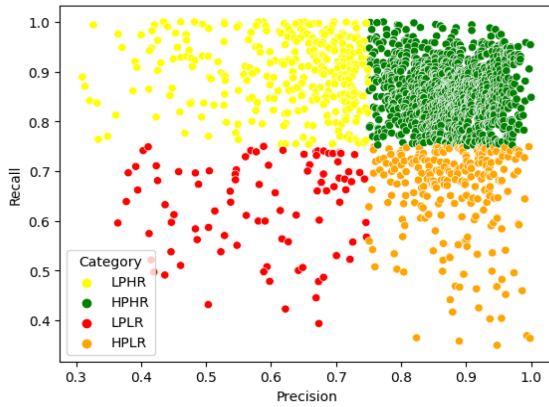


Fig. 4. Scatterplot of precision and recall for all images.

discussed method and 2) find out the shortcomings by analysis of the result. In Table I, quantitative results achieved by our model in comparison with nine distinct methods evaluated on the DeepGlobe dataset are given. As outlined, our proposed approach demonstrated better performance across recall, F-Score, and IOU than other models, and just for precision, the Res-UNet model performed better. A key point is that this model achieved good performance on both precision and recall, meaning that both predicting actual road pixels as road and not predicting the background as road have been done with high accuracy. For the road extraction task, the road is the target class while it contributes the minor part of the dataset, so the minor class is the target class. Precision and recall are the metrics that are best suited for biased datasets with minor target class. Ideally, there should be as few false predictions as possible, including both FPs and FNs. FPs and FNs are signified by the precision and recall metrics, respectively. In other words, if there is an attribute shared by samples with low precision, this attribute may be the cause of a high number of FP.

In a similar vein, any shared attribute among samples with low recall could be a potential cause of the high number of FNs.

In Fig. 4, the distribution of precision and recall of all test images has been shown. The value of 0.75 is considered a threshold for good precision and recall. With this threshold, the images are divided into the following four categories:

- 1) Images with high precision and high recall (HPHR);
- 2) Images with high precision and low recall (HPLR);
- 3) Images with low precision and high recall (LPHR);
- 4) Images with low precision and low recall (LPLR).

The goal is to figure out the causes of erroneous predictions, including FPs and FNs, by examining the model's performance on the test set. Precision and recall are affected negatively by FP and FN, respectively. We aimed to determine the causes of the model's inaccurate predictions by categorizing the results and comparing samples from various categories.

In Fig. 5, the accuracy, IOU, and F-Score distribution of the four categories are presented. It is noteworthy that while the distribution of IOU and F-Score differs for each category, accuracy is consistent across the categories and centered around the same value. The explanation for this is that accuracy mimics true predictions, whereas IOU and F-Score represent false predictions. Precision, recall, F-Score, and IOU would all be zero for a model that predicted no-road for the entire image, but it may still achieve excellent accuracy if the road area is sufficiently low. With these points in mind, comparing and analyzing image-mask pairs from different categories provides a good insight into the limiting factors of precision and recall metrics. In the following part, we will dive into a deep analysis of the four categories.

In Fig. 6, several sample images from each category are shown. For each sample, the image, mask, and prediction of the model have been displayed to be compared easily.

Let us assume that there is no masking issue in the dataset, which means all the roads have been masked as roads, and all the background areas are labeled as backgrounds. By this assumption, FP means that the model has labeled the background as roads, and FN means the model has predicted the background as roads. By analyzing the results, we noticed that there are some cases where a road is not masked as a road in the dataset, and even in a few cases, nonroad areas were masked as roads. We call these issues undermasking and overmasking, respectively. In the case of overmasking, even if the model predicts the overmasked area as road, which is really background, FNs will increase, which would cause recall to decline. Furthermore, when the dataset is overmasked, a high recall of the model means it has predicted the overmasked areas as roads, which are not really roads and just masked as roads erroneously. When the dataset is undermasked, a high precision of the model means it has predicted the undermasked areas as background which is not background in reality. It means that when there is a masking issue in the dataset, a high value of metrics does not mean that the model extracts roads properly; it means the model predicts the same result as masks, no matter if they are roads or not. Therefore, it is very important to find out which issues are related to masking issues and which are really due to the model's performance. The issues related to the dataset could not be solved by model tuning and vice versa.

As expected, most of the predictions of our model are included in Category 1, which has the best road extraction performance, and the model's predictions align more accurately with the ground truth masks compared to other categories. Although there is a slight difference between mask and prediction, it

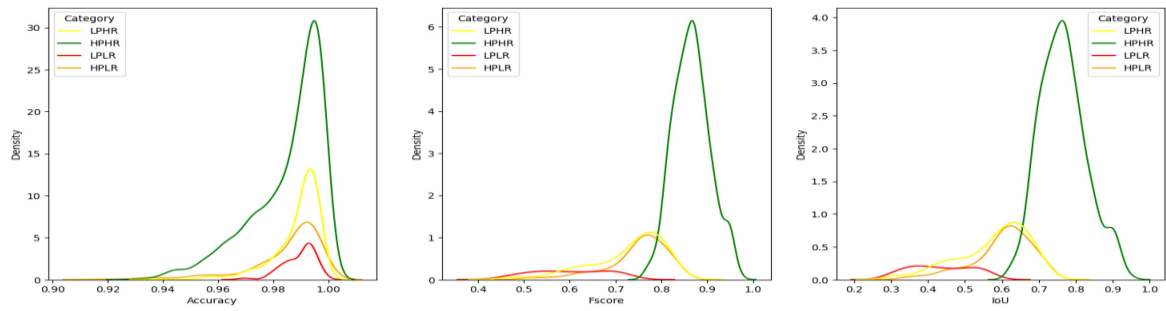


Fig. 5. Distribution of accuracy, F-score, and IOU for all images.

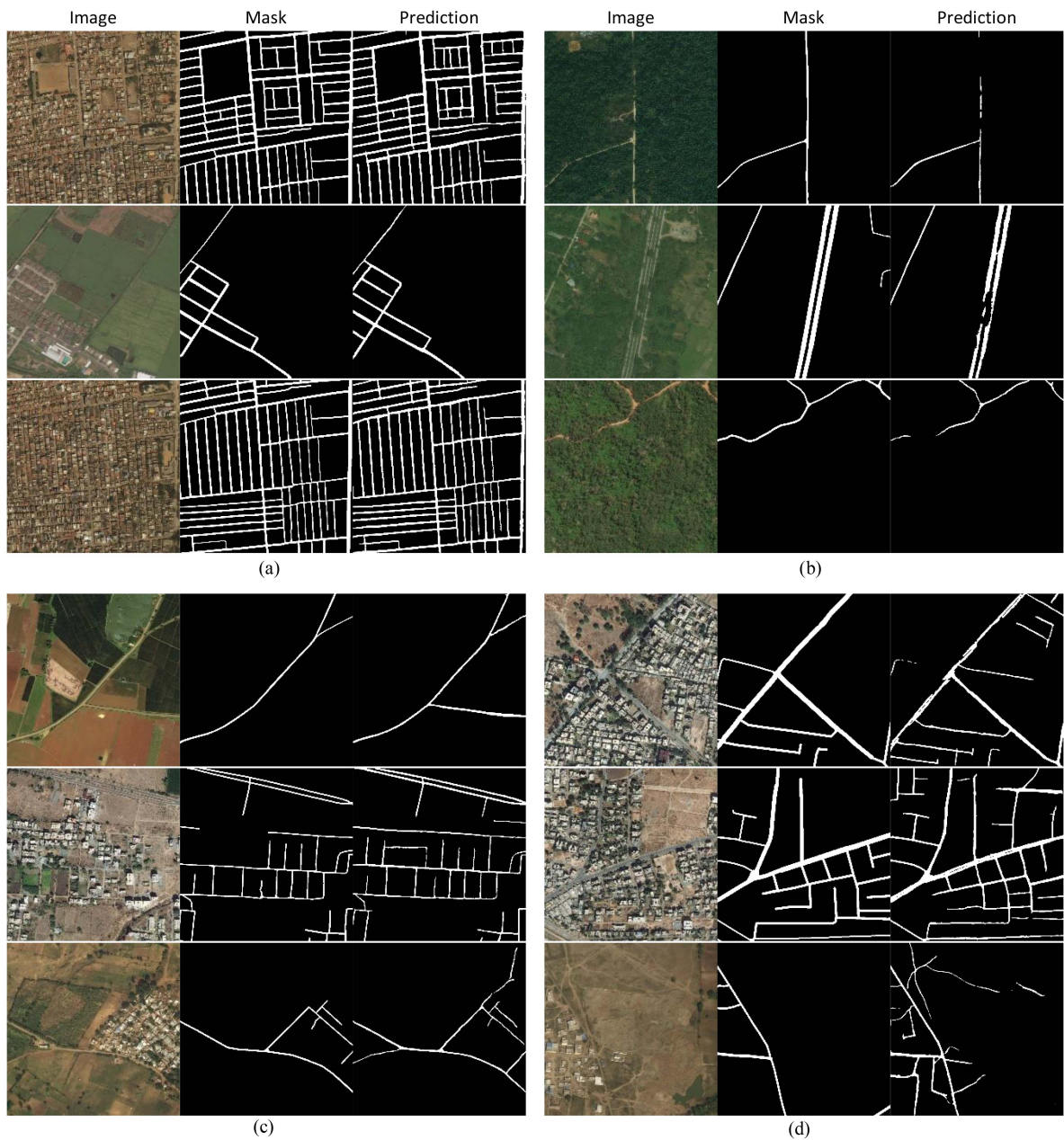


Fig. 6. Sample images of each category. (a) HPHR. (b) HPLR. (c) LPHR. (d) LPLR.

TABLE II
NUMBER OF SAMPLES IN EACH CATEGORY

Category	Count of Image
High Precision, High Recall	1071
High Precision, Low Recall	247
Low Precision, High Recall	314
Low Precision, Low Recall	94
Grand Total	1726

is far less than in other categories. In this category, we did not find any considerable masking issues, which means the shortcomings in performance are mainly related to the model itself, and analyzing these images could shed light on the model's performance and the difficulties of road extraction tasks such as occlusion. The main issue in this category is occlusion, which causes interruptions to road continuity and some erroneous island pixels detected as roads which seem that. One solution could be postprocessing the model's outputs with another model or method to clean up the erroneous pixels and check the road network continuity.

Category 2 includes images with HPLR. It means the main issue is FNs, while FPs are not considerable. The primary challenge in this category of images is that the model fails to identify certain road segments. One main reason is occlusion as shown in the sample pictures. In this category, there are cases in which seem that the prediction is completely aligned with the mask but precision is low due to the lower thickness of the predicted road compared to the mask. This issue is mostly observed for rural roads where the road edge is not clear and the narrower width prediction causes FN to increase and recall to decline. Another issue in this category which is observed in very few cases is related to the overmasked images. For such cases, although the model properly considered the overmasked area as background, it is counted as FN, which affects recall metric value. We can say that such samples cause unreal degradation of recall performance.

Category 3 includes images with LPHR. In this category, FPs are high, while FNs are not considerable. Three main issues were observed in this category. One major issue is related to the undermasked cases, for which the model has correctly detected undermasked areas as roads, but as they are counted as FPs due to the mask issue, the precision metric degrades. In Table II, the number of samples in each category has been shown. The samples in category 1 exceed the other categories significantly. It means that the model succeeded to extract the road accurately for most of the images.

It seems the dataset includes more undermasked cases than overmasked cases, as low precision due to undermasking was observed much more than low recall due to overmasking. The likelihood of encountering this category in the undermasked images is higher, and as a result, despite the model's successful performance, the metric values have diminished due to the masking issue in the dataset. In this category, there are plenty of samples for which modifying the mask would help enhance the quality of the dataset and achieve better performance for the models trained on it.

The other main issue observed in this category is related to the road-like terrestrial objects, which cause the model to be fooled

and detect them as roads. There were some test images, including road-like objects such as rivers, farmland borders, airstrips, and industrial structures detected partially as roads.

It is frequently observed that narrower and less prominent roads are not labeled in the mask, yet the model correctly identifies them. It depends on the goal of the extraction task to decide which roads should be extracted, but it is possible to mitigate the fooled model issue by creating more detailed datasets with different classes for different road types. This requires a different type of dataset and also changes the problem from binary segmentation to multiclass segmentation. With the same solution, it is possible to prevent the model from predicting road-like objects as roads. By including the road-like objects in the mask with their own class, the model, in addition to extracting more information from the image, would learn to label road-like objects in their own class. The other types of images that contributed to this category were those with very small portions of the road. The reason could be that in such cases, even if the entire road is accurately detected, identifying a small fraction of the background as a road drastically reduces precision.

In the category-4 images, the model exhibited poor performance in both precision and recall metrics. Compared to other categories, in this category, the model has identified a significant portion of the background as a road and failed to detect some parts of the actual road. Mainly, the reason for poor performance in this category is a combination of all the issues already discussed. All of the already-mentioned issues, including model failures, dataset issues including under- and overmasking, and false detection of road-like objects, were observed in this category. These cases are not only the worst predictions of the model but also the most complicated cases to be analyzed or even improved in future models, as usually the reason is a combination of multiple issues. Usually, this occurrence is more prevalent in images that include dirt roads in rural areas. Notably, the majority of images falling into this category tend to be nonurban scenes and secondary roads, while in the first category, urban images with organized road networks had a significantly higher representation.

Categorizing the images based on the precision and recall metrics and comparing samples of different categories led us to detect different reasons causing these metrics degradation. Occlusion, detecting road-like objects as road, undermasked and overmasked cases, road edge detection issues in rural roads affecting road width, and metric degradation for the samples with low portion of road area are the main issues observed in different categories and discussed. Among the issues, mask issues are different from others as they are related to the dataset affecting the metrics. In such cases, the metrics are not reliable to judge the performance of the model and its predictions. As these issues are mainly observed in categories 2 and 3, these categories are the best candidates for dataset modification.

There are different reasons that make road extraction from remote sensing images a challenging task. These challenges can be related to the shortcomings of the model or the issues of the dataset. Among the model-related factors, we can address inappropriate model selection, improper model design,

and inadequate choice of parameters and hyperparameters. Furthermore, dataset deficiencies, such as low quality, insufficient data volume for model training, limited diversity within the dataset, and skewed dataset, can have a negative impact on model performance. In addition to the mentioned factors, in supervised learning algorithms, issues related to labeling and masking of images pose significant challenges.

The overmasking and undermasking issues of the dataset were addressed as one reason for low precision and recall, respectively. As these metrics degradations were due to mask issues and not model performance, we called it fake performance degradation that does not resemble the real performance of the model.

As the images in categories 2 and 3 were mainly affected by these masking issues, they were suggested as good candidates for dataset mask modification with some considerations. For the issue of the fooled model where it predicts road-like objects as road and the issue of different types of roads detection, it was suggested that different types of datasets with multiclass targets would achieve higher performance while adding difficulty to the problem on the other hand.

IV. CONCLUSION

In our proposed model, we attempted to implement an appropriate approach for road extraction from remote sensing images using UNet architecture and techniques such as dataset augmentation and deploying attention blocks in the model. This was aimed at achieving better results in comparison to similar models and finally resulted in an accuracy of 98.33% and higher performance on Recall, IOU, F-Score. Finally, by analyzing the model's performance on the test set using various metrics, we aimed to evaluate its effectiveness. Additionally, we endeavored to identify factors influencing the model's performance in road extraction from each image in the test set. To achieve this, we focused on two crucial metrics: precision and recall. We examined four distinct categories of images, which have been discussed comprehensively in previous sections. This study also addresses the intricate challenges of extracting information from remote sensing images using deep learning. Through dataset preparation and model implementation, combining different techniques such as attention-enhanced UNet models, patch-based attention mechanisms, and rotation-based augmentation, coupled with in-depth analysis of the results, we tried to achieve better results than previous works and also provide suggestions and solutions to the inherent complexities of road extraction from aerial images. Looking ahead, these insights will pave the way for refining datasets, enhancing model architectures, and propelling advancements in the realm of remote sensing applications.

REFERENCES

- [1] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602312.
- [2] M. Mohammadi and A. Sharifi, "Evaluation of convolutional neural networks for urban mapping using satellite images," *J. Ind. Soc. Remote Sens.*, vol. 49, no. 9, pp. 2125–2131, 2021, doi: [10.1007/s12524-021-01382-x](https://doi.org/10.1007/s12524-021-01382-x).
- [3] A. Kosari, A. Sharifi, A. Ahmadi, and M. Khoshshima, "Remote sensing satellite's attitude control system: Rapid performance sizing for passive scan imaging mode," *Aircr. Eng. Aerosp. Technol.*, vol. 92, no. 7, pp. 1073–1083, 2020, doi: [10.1108/AEAT-02-2020-0030](https://doi.org/10.1108/AEAT-02-2020-0030).
- [4] A. Zamani, A. Sharifi, S. Felegari, A. Tariq, and N. Zhao, "Agro climatic zoning of saffron culture in Miyaneh city by using WLC method and remote sensing data," *Agriculture*, vol. 12, no. 1, 2022, Art. no. 118, doi: [10.3390/agriculture12010118](https://doi.org/10.3390/agriculture12010118).
- [5] M. Mohammadi, A. Sharifi, M. Hosseingholizadeh, and A. Tariq, "Detection of oil pollution using SAR and optical remote sensing imagery: A case study of the Persian gulf," *J. Ind. Soc. Remote Sens.*, vol. 49, no. 10, pp. 2377–2385, 2021, doi: [10.1007/s12524-021-01399-2](https://doi.org/10.1007/s12524-021-01399-2).
- [6] A. Sharifi, "Flood mapping using relevance vector machine and SAR data: A case study from Aqqala, Iran," *J. Ind. Soc. Remote Sens.*, vol. 48, no. 9, pp. 1289–1296, 2020, doi: [10.1007/s12524-020-01155-y](https://doi.org/10.1007/s12524-020-01155-y).
- [7] A. Tariq et al., "Flash flood susceptibility assessment and zonation by integrating analytic hierarchy process and frequency ratio model with diverse spatial data," *Water*, vol. 14, no. 19, 2022, Art. no. 3069, doi: [10.3390/w14193069](https://doi.org/10.3390/w14193069).
- [8] A. Sharifi, "Development of a method for flood detection based on Sentinel-1 images and classifier algorithms," *Water Environ. J.*, vol. 35, no. 3, pp. 924–929, 2021, doi: [10.1111/wej.12681](https://doi.org/10.1111/wej.12681).
- [9] Z. Chen et al., "Road extraction in remote sensing data: A survey," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, 2022, Art. no. 102833.
- [10] P. Liu, Q. Wang, G. Yang, L. Li, and H. Zhang, "Survey of road extraction methods in remote sensing images based on deep learning," *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.*, vol. 90, pp. 135–159, 2022.
- [11] A. Sharifi, J. Amini, J. T. Sri Sumantyo, and R. Tateishi, "Speckle reduction of PolSAR images in forest regions using fast ICA algorithm," *J. Ind. Soc. Remote Sens.*, vol. 43, no. 2, pp. 339–346, 2015, doi: [10.1007/s12524-014-0423-3](https://doi.org/10.1007/s12524-014-0423-3).
- [12] A. Sharifi and J. Amini, "Forest biomass estimation using synthetic aperture radar polarimetric features," *J. Appl. Remote Sens.*, vol. 9, no. 1, 2015, Art. no. 097695, doi: [10.1117/1.jrs.9.097695](https://doi.org/10.1117/1.jrs.9.097695).
- [13] C. Poullis and S. You, "Delineation and geometric modeling of road networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 2, pp. 165–181, 2010.
- [14] N. Farmonov et al., "Crop type classification by DESIS hyperspectral imagery and machine learning algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1576–1588, 2023, doi: [10.1109/JSTARS.2023.3239756](https://doi.org/10.1109/JSTARS.2023.3239756).
- [15] S. M. M. Nejad, D. Abbasi-Moghadam, A. Sharifi, N. Farmonov, K. Amankulova, and M. Laszlj, "Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 254–266, 2023, doi: [10.1109/JSTARS.2023.3223423](https://doi.org/10.1109/JSTARS.2023.3223423).
- [16] M. Esmaili, D. Abbasi-Moghadam, A. Sharifi, A. Tariq, and Q. Li, "Hyperspectral image band selection based on CNN embedded GA (CN-NeGA)," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1927–1950, 2023, doi: [10.1109/JSTARS.2023.3242310](https://doi.org/10.1109/JSTARS.2023.3242310).
- [17] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sens.*, vol. 12, no. 9, 2020, Art. no. 1444.
- [18] M. Mokhtarzade and M. V. Zojc, "Road detection from high-resolution satellite images using artificial neural networks," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, pp. 32–40, 2007.
- [19] B. Tu, X. Yang, W. He, J. Li, and A. Plaza, "Hyperspectral anomaly detection using reconstruction fusion of quaternion frequency domain analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2022.3227167](https://doi.org/10.1109/TNNLS.2022.3227167).
- [20] B. Tu, Q. Ren, Q. Li, W. He, and W. He, "Hyperspectral image classification using a superpixel-pixel-subpixel multilevel network," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5013616.
- [21] M. Saeedimoghaddam and T. F. Stepinski, "Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks," *Int. J. Geographical Inf. Sci.*, vol. 34, pp. 947–968, 2020.
- [22] F. Bastani et al., "RoadTracer: Automatic extraction of road networks from aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4720–4728.
- [23] X. Lu et al., "Multi-scale and multi-task deep learning framework for automatic road extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9362–9377, Nov. 2019.

- [24] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 1591–1594.
- [25] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 210–223.
- [26] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, 2017.
- [27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [28] N. Varia, A. Dokania, and J. Senthilnath, "DeepExt: A convolution neural network for road extraction using RGB images captured by UAV," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2018, pp. 1890–1895.
- [29] R. Kestur, S. Farooq, R. Abdal, E. Mehraj, O. S. Narasipura, and M. Mudigere, "UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle," *J. Appl. Remote Sens.*, vol. 12, 2018, Art. no. 016020.
- [30] T. Panboonyuen, P. Vateekul, K. Jitkajornwanich, and S. Lawawirojwong, "An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery," in *Proc. Int. Conf. Comput. Inf. Technol.*, 2017, pp. 191–201.
- [31] Z. Hong, D. Ming, K. Zhou, Y. Guo, and T. Lu, "Road extraction from a high spatial resolution remote sensing image based on richer convolutional features," *IEEE Access*, vol. 6, pp. 46988–47000, 2018.
- [32] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascade end-to-end convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "UNet: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [34] R. Azad et al., "Medical image segmentation review: The success of UNet," 2022, *arXiv:2211.14830*.
- [35] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [36] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [38] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 182–186.
- [39] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
- [40] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," *Inf. Sci.*, vol. 535, pp. 156–171, 2020.
- [41] Y. Hou, Z. Liu, T. Zhang, and Y. Li, "C-UNet: Complement UNet for remote sensing road extraction," *Sensors*, vol. 21, no. 6, 2021, Art. no. 2153.
- [42] M. Yang, Y. Yuan, and G. Liu, "SDUNet: Road extraction via spatial enhanced and densely connected UNet," *Pattern Recognit.*, vol. 126, 2022, Art. no. 108549.
- [43] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.
- [44] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," 2018, *arXiv:1804.02391*.
- [45] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [46] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1461.
- [47] Y. Ren, Y. Yu, and H. Guan, "DA-CapsUNet: A dual-attention capsule UNet for road extraction from remote sensing imagery," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2866.
- [48] J. Wan, Z. Xie, Y. Xu, S. Chen, and Q. Qiu, "DA-RoadNet: A dual-attention network for road extraction from high resolution satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6302–6315, 2021.
- [49] J. Li, Y. Liu, Y. Zhang, and Y. Zhang, "Cascaded attention DenseUNet (CADUNet) for road extraction from very-high-resolution images," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 5, 2021, Art. no. 329.
- [50] S. Li et al., "Cascaded residual attention enhanced road extraction from remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 1, 2022, Art. no. 9.
- [51] S. Shao, L. Xiao, L. Lin, C. Ren, and J. Tian, "Road extraction convolutional neural network with embedded attention mechanism for remote sensing imagery," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2061.
- [52] I. Demir et al., "DeepGlobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–17209.
- [53] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7570–7588, 2021.
- [54] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2018, Dec. 2018, pp. 6389–6399.
- [55] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "UNet and its variants for medical image segmentation: Theory and applications," 2020, *arXiv:2011.01118*.
- [56] D. Soydaner, "Attention mechanism in neural networks: Where it comes and where it goes," *Neural Comput. Appl.*, vol. 34, no. 16, pp. 13371–13385, 2022.
- [57] H. Xu, H. He, Y. Zhang, L. Ma, and J. Li, "A comparative study of loss functions for road segmentation in remotely sensed road datasets," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 116, 2023, Art. no. 103159.
- [58] J. Terven, D. M. Cordova-Esparza, A. Ramirez-Pedraza, and E. A. Chavez-Urbiola, "Loss functions and metrics in deep learning. A review," 2023, *arXiv:2307.02694*.
- [59] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.



Arezou Akhtarmanesh received the B.S. and M.Sc. degrees in telecommunication electrical engineering from Shahid Bahonar University, Kerman, Iran, in 2015 and 2023, respectively.

Her research interests include wireless communications, image processing, and machine learning.



Dariush Abbasi-Moghadam received the B.S. degree in electrical engineering from Shahid Bahonar University, Kerman, Iran, in 1998 and the M.S. and Ph.D. degrees from the Iran University of Science and Technology, Tehran, Iran, in 2001 and 2011, respectively, all in electrical engineering.

He was primarily with the Advanced Electronic Research Center—Iran from 2001 to 2003 and worked on the design and analysis of satellite communication systems. In September 2004, he joined the Iranian Telecommunications Company, Tehran, as a Research Engineer. He is currently with the Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran, as an Associate Professor. His research interests include the areas of wireless communications, satellite communication systems, remote sensing, and signal processing.



Alireza Sharifi was born in Tehran, Iran, in 1981. He received the M.Sc. and Ph.D. degrees in remote sensing engineering from the University of Tehran, Tehran, Iran, in 2008 and 2015, respectively.

He is currently an Associate Professor of remote sensing with the Faculty of Civil Engineering, Shahid Rajaei Teacher Training University, Tehran, Iran. In particular, he is involved in the GEOAI program for food security and environmental monitoring.



Aqil Tariq received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2021.

He is currently with the Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Mississippi State, MS, USA. He is the author of more than 130 journal articles and conference proceedings papers. His research interests include photogrammetry, precision agriculture, 3-D geoinformation, urban analytics, spatial analysis to examine land use/land cover, geospatial data science, agriculture monitoring, natural hazards (forest fire, landslide, droughts, and flood) forest monitoring, forest cover dynamics, spatial statistics, multicriteria algorithms, ecosystem sustainability, hazards risk reduction, statistical analysis, and modeling using Python, R, and MATLAB.



Mohsen Hazrati Yadvouri was born in the Qashqai tribe in the south of Iran. He received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 2010, and the M.Sc. degree in nanoelectronic devices from the Sharif University of Technology, Tehran, Iran; however, he transitioned to the cellular mobile telecommunication industry before completing the master's degree.

He is currently a RAN engineer in the cellular network industry and has a keen interest in machine learning.



Linlin Lu received the Ph.D. degree in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 2009.

Since 2013, he has been an Associate Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences. She is the author of more than 120 journal articles and conference proceeding papers. Her research interests include image information detection, image classification, and time-series

analysis applied to urban environment and urban sustainability.

Dr. Lu was appointed as a Member of Sino-EU Panel on Land and Soil (SE-PLS) in 2017. She is a Member of Group on Earth Observations (GEO) Global Urban Observation and Information Initiative and Human Planet Initiative. She currently co-chairs the Urban Environment Working Group in the Digital Belt and Road program.