IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 9, 2020, accepted November 11, 2020, date of publication November 16, 2020,
date of current version December 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038225

Building Footprint Extraction from High Resolution Aerial Images Using Generative Adversarial Network (GAN) Architecture

ABOLFAZL ABDOLLAHI[1], BISWAJEET PRADHAN[1,2,3], (Senior Member, IEEE), SHILPA GITE[4], AND ABDULLAH ALAMRI[5]

[1]Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia
[2]Department of Energy and Mineral Resources Engineering, Sejong University, Seoul 05006, South Korea
[3]Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, Selangor 43600, Malaysia
[4]Computer Science and Information Technology Department, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune 412115, India
[5]Department of Geology and Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

Corresponding author: Biswajeet Pradhan (Biswajeet.Pradhan@uts.edu.au)

This work was funded in part by the Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT; in part by the University of Technology Sydney (UTS); and in part by the Researchers Supporting Project number RSP-2020/14, King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Building extraction with high accuracy using semantic segmentation from high-resolution remotely sensed imagery has a wide range of applications like urban planning, updating of geospatial database, and disaster management. However, automatic building extraction with non-noisy segmentation map and obtaining accurate boundary information is a big challenge for most of the popular deep learning methods due to the existence of some barriers like cars, vegetation cover and shadow of trees in the high-resolution remote sensing imagery. Thus, we introduce an end-to-end convolutional neural network called Generative Adversarial Network (GAN) in this study to tackle these issues. In the generative model, we utilized SegNet model with Bi-directional Convolutional LSTM (BConvLSTM) to generate the segmentation map from Massachusetts building dataset containing high-resolution aerial imagery. BConvLSTM combines encoded features (containing of more local information) and decoded features (containing of more semantic information) to improve the performance of the model even with the presence of complex backgrounds and barriers. The adversarial training method enforces long-range spatial label vicinity to tackle with the issue of covering building objects with the existing occlusions such as trees, cars and shadows and achieve high-quality building segmentation outcomes under the complex areas. The quantitative results obtained by the proposed technique with an average F1-score of 96.81% show that the suggested approach could achieve better results through detecting and adjusting the difference between the segmentation model output and the reference map compared to other state-of-the-art approaches such as autoencoder method with 91.36%, SegNet+BConvLSTM with 95.96%, FCN-CRFs with 95.36%% SegNet with 94.77%, and GAN-SCA model with 96.36% accuracy.

INDEX TERMS Building extraction, GAN, remote sensing, SegNet.

I. INTRODUCTION

One of the main stages in geospatial information system (GIS) applications, such as change detection, urban land use analysis, geospatial database updating, and infrastructure planning, is the automatic extraction of features, like building objects from high-resolution remote sensing imagery. Given that remote sensing imagery normally involves data in the format of incoherent areas with low inter-class and high intra-class differences, classifying pixels into semantic objects is one of the most significant and challenging obstacles in the high-resolution satellite and aerial imagery

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

of urban domains [1]. Other challenges exist in these images, such as shadows, overlapping, and interlacing sheltering [2]. These challenges are even more pronounced in the extraction of urban features, like building objects [3]. Most of the current approaches that usually depend on a collection of pre-defined characteristics are restricted by such heterogeneities in remote sensing imagery [4]. Therefore, designing a robust approach that could achieve high-precision features extraction on high-resolution remote sensing imagery is very difficult [5]. In recent years, effective state-of-the-art proficiency has been obtained using deep learning models for semantic classification [6], [7] not only by some methods that have been suggested in the field of computer vision, such as integrated convolutional neural network (CNN) models with segmentation [8], the deep parsing model [9], patch network [10], the combination of conditional random fields (CRFs) with CNN [11], DeepLab [12], decoupled networks [13], deconvolutional models [14], and SegNet [15], but also in the field of remote sensing [16], [17].

As remote sensing imagery, such as high-spatial, high-spectral resolution imagery and multi-spectral images, are intrinsically and typically big data, deep convolutional neural networks (DCNN) are considered for leaning hierarchical features and extracting semantic features and information from these images [18]–[20].

DCNNs contain several convolutional and inter-connected layers that can learn a hierarchical feature representation from an image and encode spatial and spectral information efficiently on the basis of raw pixel-data input without requiring any preprocessing [21]. Consequently, DCNNs are rapidly becoming prominent methods in remote sensing applications that identify characteristics in various representation levels [22]. Generally, two major methods are used in the architectures of CNN models for semantic segmentation: 1) end-to-end methods (pixel based) [22] and 2) patch-based approaches [23].

In pixel-based approaches, encoder–decoder or fully convolutional network (FCN) architectures are normally utilized by performing interpolation and up-sampling to identify the fine detail of the raw input imagery [24]. Meanwhile, patch-based approaches utilize small patches of image to start the training process and then utilize the sliding window technique to forecast every class pixel [23]. This method is normally utilized for detecting large objects of urban areas [3].

In this study, we also present a pixel-based deep leaning method named Generative Adversarial Network (GAN) [25] to extract building features from Massachusetts building datasets that include high-resolution remote sensing aerial imagery. The GAN method utilizes two generative and discriminative training models. The data dispensation is captured by the generative part, while the likelihood that an example relates to fake (generated) or real one (from the domain) is estimated by the discriminator part [26].

This research aims not only to obtain a map of segmentation with many details that explain the boundary information, but also a non-noisy map of segmentation considering high spatial contiguity. Therefore, we utilized an encode–decoder model (i.e., SegNet with BConvLSTMs) for the generator part of the proposed GAN model to generate a high-quality segmentation map with a similar resolution as the input image. For combining the encoded and decoded features, we utilized a set of BConvLSTM in the decoding part of the SegNet model. The decoded features contain semantic information about the input data, whereas the encoded features include more local information and have higher resolution of the input data. Thus, we used BConvLSTM to mix the encoded and decoded features rather than using a simple concatenation. This is because a set of feature maps rich in both semantic and local information might be resulted by the influence of these two encoded and decoded features. Every ConvLSTM case matches to one type of features (encoded features) is capable to encode related information about the other type of features (decoded features). This is because a set of convolutional filters are performed on every type of features in the BConvLSTM. The hyperbolic tangent functions coupled with convolutional filters assist the model to learn structures of data.

The introduced GAN model obtained more constant building outcomes while enforcing long-range spatial label vicinity and improved the performance. The rest of this paper is organized as follows. Section II presents a review of relevant building extraction studies. Section III describes the methodology of the proposed GAN model for building extraction. The visual and quantitative results obtained by the proposed GAN method and other comparison approaches are presented in Section IV. and Section V provides the conclusion.

II. RELATED WORKS

In this part, the relevant methods of CNNs for building extraction from high-resolution remotely sensed imagery and their major contributions are discussed, and the contribution of this research is also highlighted at the end of this part.

Many studies related to building extraction using deep convolutional neural networks have been conducted. The patch-based CNN model was first suggested for building extraction by [27]. Instead of a batch-based CNN model, [28] applied an object-based segmentation CNN approach with a similar architecture for building extraction from orthorectified images with a 12-cm spatial resolution. A single-architecture CNN model was applied by [29], [30] on the Massachusetts dataset for road and building extraction. They depicted the ability of a single CNN architecture in the simultaneous semantic segmentation of multiple features, such as roads, background, and buildings. To improve the performance of the CNN method, [29] utilized dropout optimization with an extra MAXOUT layer rather than the activation function of rectified linear unit (RELU), while [30] applied a function called channel-wise inhibited softmax and utilized an averaging model with the spatial shift method to repress the background influence and the semantic segmentation. The abovementioned studies produced good experimental results for building features segmentation from high-resolution

remote sensing imagery. However, the patch-based CNN model produces output patches with discontinuity boundaries. In addition, the CNN model performs well on single houses extraction but not on complicated urban areas and complex buildings [31], confirming that only dependent pixels inside patches can be classified by the patch-based CNN architecture and not the independent ones [32]. Furthermore, this method is difficult to use on remote sensing imagery as tiny patches cannot obtain the information of entire single buildings, covering only fragmented buildings [33].

ImageNet framework was applied by [34] for supervised building extraction to address the limitations of the patch-based CNN model. However, the spatial information at a finer resolution that is vital for dense anticipation was discarded using a fully convolutional layer at the end. Moreover, [34] used a patch-based sliding window for training and testing procedures, which is time-consuming [31]. A pixel-based FCN was proposed by [32] to produce a dense anticipation. The original size of the input image was upsampled by adding a deconvolutional layer. The discontinuity problem was solved by the process, and the accuracy was improved because of the short time required and the uncomplicated learning procedure. An encode–decoder convolutional neural network was applied by [31] on multisource remote sensing datasets. They used a Massachusetts dataset for pre-training and combined the NRG and RGB bands together for accurate building extraction. By adding a deconvolutional layer, the memory requirement and complexity of the model for training increased excessively, but the results achieved by the proposed DeCNN model improved. Bittner *et al.* [35] applied a technique based on the digital surface model and the FCN on different datasets for building extraction. The proposed FCN model was fine-tuned and constructed based on the VGG-16 model [36]. Eventually, the CRFs were used to achieve a binary mask of buildings. However, the fully connected CRFs used at the end of the original FCN model did not provide a satisfying outcome. The VGG-Net [36] and Alex-Net [6] models have been used to extract the local and global contexts of building features, respectively. The Alex-Net model can detect global information due to its large filter size, while the VGG-Net model, which has a small filter size, can detect local information. Therefore, this model is very complex because of the combination of two single CNN models.

The deep learning methods have been used recently for semantic segmentation in extracting building features which can be regarded as a semantic segmentation issue [37]. The problem of pixel-wise labelling can be addressed by semantic segmentation techniques like FCN suggested by [38], which are utilized through computer vision association. Chen *et al.* [39] applied a DCNN model for semantic segmentation based on either the ResNet-101 [40] or VGG-16 model. In the proposed network, atreus convolutional layers were used to increase the feature resolution, and the CRFs method was used to purify the segmentation outcomes. However, the delicate object borders were not obtained by the proposed network. Noh *et al.* [14] introduced a new method captured from the VGG-16 architecture for semantic segmentation. To address the issue of pixel-wise labelling, the deconvolutional model consisted of unspooling and deconvolutional layers. Shrestha and Vanneschi [37] proposed a method named improved fully convolutional framework for building semantic segmentation. To improve the efficiency of the proposed model, they used CRFs method to sharpen the building borders. Xu *et al.* [41] applied the Res-U-Net model for building extraction from the Potsdam and Vaihingen datasets. They used a guided filter in the post-processing step to remove salt-and-paper noise and improve the results. However, the model could not classify some irregular and blurry boundaries and could not detect buildings surrounded by trees accurately. Aung *et al.* [42] applied conditional generative adversarial network (CGAN) to extract building footprint from GeoEye images of Yangon city, Myanmar. Evaluation metrics such as completeness, correctness and f1 are calculated and based on the outcomes, the proposed method indicated promising performance in building detection. Shao *et al.* [43] implemented a new deep-based network for building extraction from Massachusetts building images based on Building Residual Refine Network (BRRNet), which includes two sections as prediction and the residual refinement modules. During training, they used dice loss function to reduce the data imbalance issue, and the results showed the superiority of the proposed model in building extraction.

Although the abovementioned techniques have achieved certain accomplishments in addressing the issue of building extraction, they have certain deficiencies [31]. For example, in heterogeneous parts with some obstacles, such as shadow, trees, and cars, most of these methods exhibit poor performance in building segmentation [34]. Therefore, a new approach based on GAN model that induces long-range spatial label cognitive is applied in this study for building detection to mitigate the aforementioned weaknesses. Furthermore, a homogenous building outcome can be obtained by the suggested technique, even under the obstacles of shadows and trees or in the heterogeneous parts.

III. METHODOLOGY

The concept of the GAN model is first explained in this part. Then, the application of GAN for semantic segmentation is discussed. To apply the proposed model in building extraction, training and test images were prepared from the original Massachusetts building dataset. Then, the training images with corresponding labels were used to train the GAN model. Subsequently, the trained GAN model was implemented on test images for detecting building features and generating a segmentation map. Finally, evaluation metrics were utilized to assess the effectiveness of the GAN model, and visual and quantitative outcomes are presented. Figure 1 presents the overall methodology of the current work.

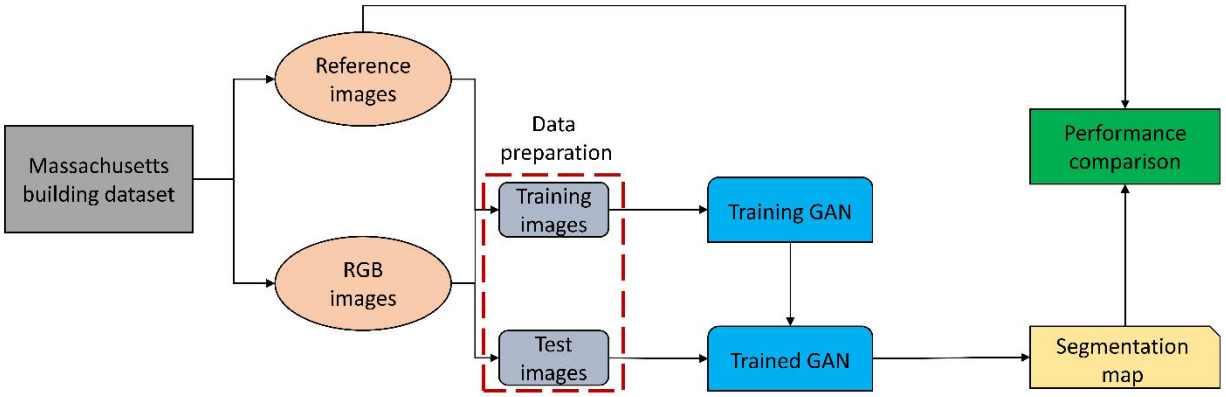

FIGURE 1. The overall architecture of the suggested GAN network.

A. GAN ARCHITECTURE

The GAN network comprises two main parts, namely, the generative and discriminator terms [44]. The data possibility dispensation was stimulated by the generative model, while the discriminator term was utilized to distinguish whether a specimen is obtained from the true map or the generative model. An adversarial train was formed from the generative and discriminator models to obtain the final outcomes (Figure 2). The possibility of data dispensation p_g over dataset x can be learnt by the generative model. A previous noise parameter ($p_z(z)$) was first defined, and $G(z; \theta_g)$ was used to represent a mapping from $p_z(z)$ to the output data, where output data y was produced by the G function, which was a recognizable function with θ_g variable. The possibility that x comes from the ground truth image instead of being stimulated by G is represented by the discriminator model defined as $D(x; \theta_d)$, where θ_d is the variable in the discriminator term [45].

$$\begin{aligned}\min_G \max_D L_{cGAN}(G, D) \\ = E_{y\sim p_{data}(y)}[\log D(y)] \\ + E_{x\sim p_{data}(x), z\sim p_z(z)}[\log(1 - D(G(x, z))] \end{aligned} \quad (1)$$

log (1 - D(G(x; z)) was minimized by training G through the second term, while an accurate forecast of output representation y was produced by maximizing the possibility through the first term. Although the neighborhood representation on the image contributes a similar label to a specific extent, the output samples were considered conditionally autonomous from each other by the GAN. Hence, the local area representation can be regarded as dependent. For learning the mapping from random noise parameters ($p_z(z)$) and recognized image x to output y, the conditional GAN defined by G(x; z; θ_g) and traditional loss like L2 distance [26] was suggested. Equation 2 shows the conditional GAN as

$$G^* = L_{L2}(G) + \arg\min_G \max_D L_{cGAN}(G, D) \quad (2)$$

B. SEMANTIC SEGMENTATION BY GAN MODEL

The GAN model was adapted for semantic segmentation. The output map of the proposed model for building segmentation can be denoted as y of size $H \times W \times C$ if we consider the input image as a RGB image of size $H \times W \times 3$, where C is the number of classes, W is the weight, and H is the height of the image. To magnify the likelihood of assigning the true label for every pixel in the image using the segmentation model, the first term in the GAN model was utilized, and the right label was discriminated from the output of the segmentation model by the second term. Furthermore, this term penalizes the inconformity in the statistics of the higher-order label between the right and anticipated labels [46].

Particularly in the encoder–decoder architecture of the segmentation model, the hierarchical layer of encoders and decoders indicates that the decoder layers pursue the encoders, and every decoder layer has a corresponding encoder. Figure 3 presents the suggested segmentation model workflow. For this purpose, an encoder–decoder SegNet architecture with BConvLSTM was used for the generative section to make the segmentation map with a similar input image resolution and the RELU function was used as an activation function. In addition, max-pooling indices were used to combine semantically comparable features into one as well as carry out spatial subsampling [41]. The trained max-pooling indices from the encoder layers were used by the corresponding decoder layers. The adversarial framework is described in detail below.

In our framework, classifier D was represented by binary-cross entropy [47], assuming N training images in dataset x_n and matching mask images y_n. The scalar possibility was presented by $a(x, y)in[0, 1]$ that y is the right label of x (Equation 3).

$$\begin{aligned} G_* &= L_{L2}(G) + \arg\min_G \max_D L_{cGAN}(G, D) \\ &= \sum_{n=1}^{N} l_{mce}(s(x_n), y_n) - \lambda[l_{bce}(a(x_n, y_n), 1) \\ &\quad + l_{bce}(a(x_n, s(x_n)), 0)] \end{aligned} \quad (3)$$

The binary-cross entropy loss function is denoted by $l_{bce}(z*, z) = -[z \ln z^* + (1 - z) \ln(1 - z*)]$ in the second term and utilized to train the discriminator to execute the right choices, while $l_{mce}(y*, y) = -\sum_{i=1}^{H\times W} \sum_{c=1}^{C} y_{ic} y_{ic}^*$ in the first

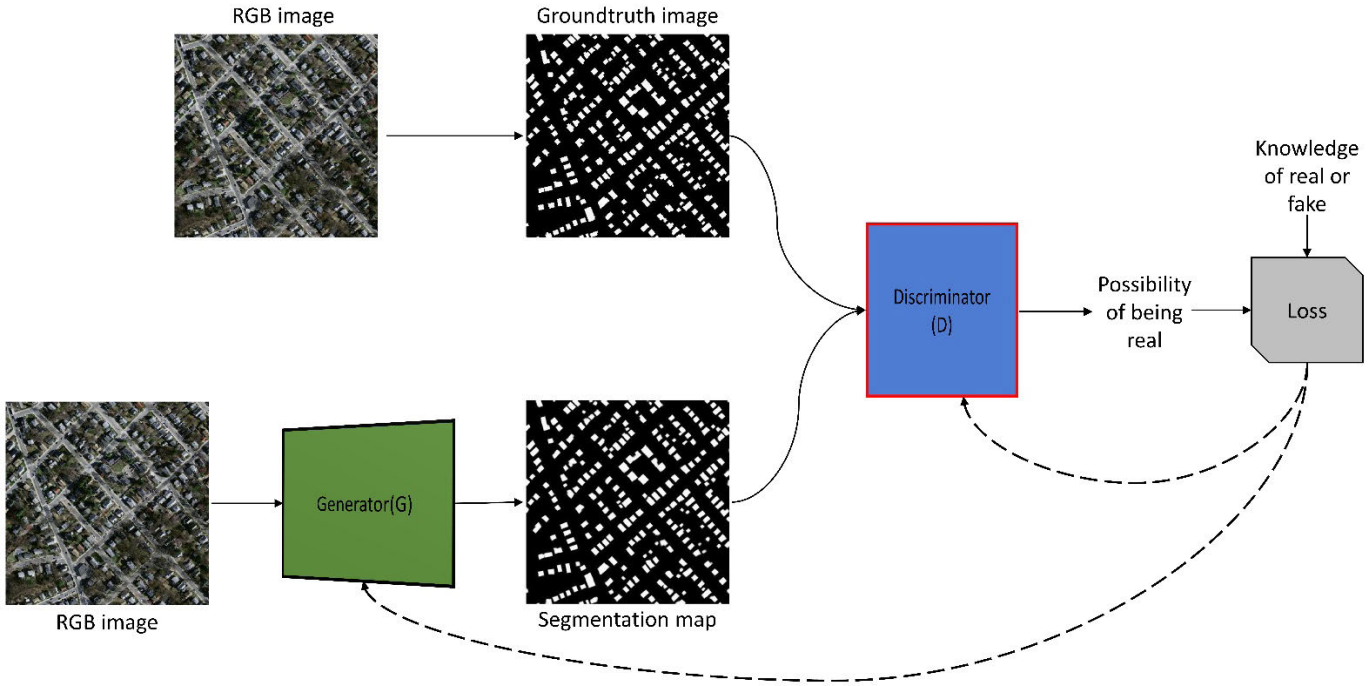

FIGURE 2. General architecture of GAN network that consist of generative and discriminative models.

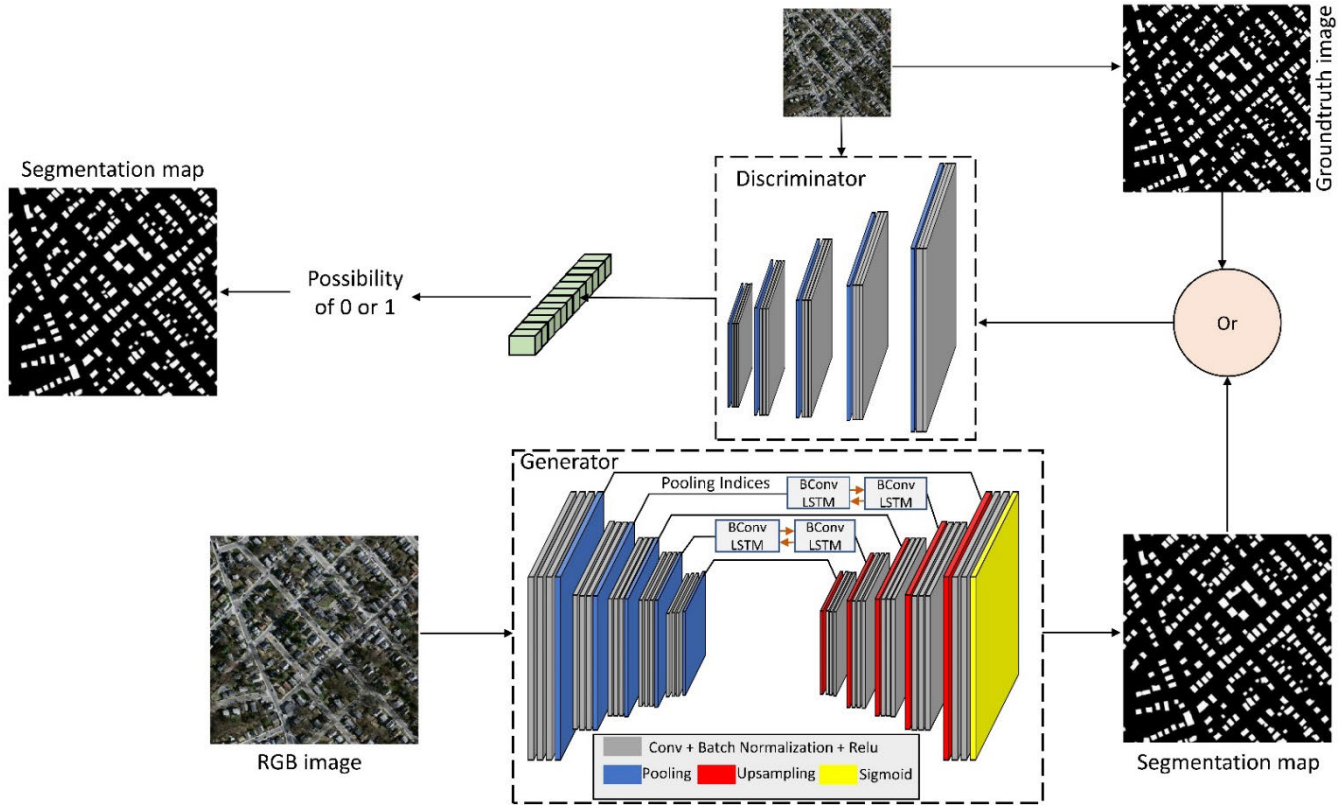

FIGURE 3. Generating label for RGB image by training GAN model. Segmentation map is produced by the generative model, while the discriminative model attempts to predict the possibility that the label map comes from the created map or the true map by taking the reference or segmentation map as an input and combining it with the RGB image.

term indicates the loss of entropy between ground truth image y_n and predicted label $s(x_n)$ and utilized for segmentation model optimization to produce representations relative to the ground truth images. We maximized the variables in the adversarial model while minimizing the loss in the segmentation model parameters. For optimizing the objective function,

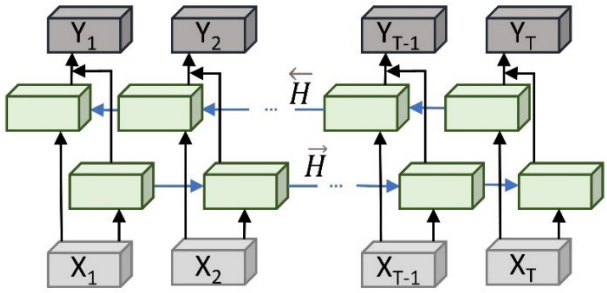

FIGURE 4. Illustration of BConvLSTM with forward and backward directions.

the alternative optimization was utilized. By fixing the generator part, the optimization of the discriminator part was first performed as just the second term consists of the adversarial procedure. Therefore, one step optimization was performed on the generative term after optimizing the discriminator term with one step using the gradient descent step. In the current work, one of the most prevalent optimizers (Adam) with learning rate of 0.0001 was utilized to decrease the losses and update the parameters, such as weights and biases. When the discriminator term is optimized [24] that is listed below:

$$\min \sum_{n=1}^{N} l_{bce}(a(x_n, y_n), 1) + l_{bce}(a(x_n, s(x_n)), 0) \qquad (4)$$

this objective function was utilized for the above binary classification loss minimization. This term could be optimized as a CNN network in which the CNN network input includes the RGB image and the corresponding reference map. For the reference map, two possible options can be taken: one is identified label $s(x_n)$ and the other is true label y_n. Given that the two different resources of inputs present various low-level representations, two branches were utilized in the adversarial framework for processing the RGB image and the corresponding reference map. Both were convolved to 32 channels to balance the influence of these two types of signals. Sigmoid activation function was utilized at the final layer of the adversarial network to produce 1 or 0. Then, the generator term was optimized after fixing the parameter in the adversarial term and optimizing this term [42].

C. BI-DIRECTIONAL CONVOLUTIONAL LSTM

Bi-directional convolutional LSTM process the input image into two forward and backward directions using two ConvLSTMs. Then, it decides for the current input by tackling the dependencies of data in both paths. The dependencies of the forward path are only processed in a standard ConvLSTM. Thus, it might be efficient to consider backward dependencies and all the information in a sequence as well. It has been confirmed that the performance of model enhances by investigating the temporal aspects of both forward and backward paths [48]. Therefore, in the decoding part of the proposed SegNet network, we employed BConvLSTM [49] to mix encoded and decoded features (Figure 4). For backward and forward states, we have two sets of parameters. This is because every forward and backward state of ConvLSTM can be taken into account as a standard one. The BConvLSTM output [49] was calculated as follows:

$$Y_t = \tanh(w_y^{\vec{H}} * \vec{H}_t + w_y^{\overleftarrow{H}} * \overleftarrow{H}_t + b) \qquad (5)$$

where $Y_t \in R^{F_l \times W_l \times H_l}$ demonstrates the last output considering bidirectional information, b is the bias and $\overleftarrow{H}$ and $\vec{H}$ denote the concealed tensors for backward and forward paths, respectively.

IV. EXPERIMENTAL RESULTS AND EVALUATION

We first describe the dataset and then explain the measurement factors for calculating the performance of the comparison approaches and the proposed method. Finally, the visual building extraction outcomes and the quantitative comparisons of the suggested technique and other state-of-the-art comparison techniques in building extraction are presented.

A. DATASET EXPLANATION

The GAN model for building extraction was applied on the Massachusetts building dataset [27]. The dataset includes 137 training images and 4 validation and 10 test images with a spatial dimension of 1500 × 1500 with the spatial resolution of 0.5 m. In this study, the main images were split into 384 × 384 dimension because of computational issues. Then, 1545 good quality images with complete information and corresponding labels were selected that 1527 images are used for training and validation and 18 images are used for testing. We used different data augmentation techniques such as rotation, zooming, vertical flip and horizontal flip to expand the size of the dataset [21]. Moreover, to overcome the over-fitting issue, a dropout of 0.5 was added to deeper convolutional layers. The proposed model was trained with a batch size 1 for 100 epochs, and the entire process of the suggested approach for building extraction from remote sensing imagery was performed on a GPU Nvidia Quadro RTX 6000 with a memory of 24 GB and a computing capability of 7.5 under the framework of Keras with a Tensorflow back-end. Figure 5 depicts some samples in the building dataset of different scenes.

B. PERFORMANCE EVALUATION

For calculating the performance of the suggested model in building extraction from high-resolution remote sensing images, we used four main evaluation factors, namely, recall, precision, f1 score, overall accuracy and intersection over union (IOU) [50]. Recall (6) is expressed as the proportion of building pixels that are anticipated correctly among all the real building pixels. The precision (7) metric is explained as the percentage of pixels determined precisely among the predicted building pixels. The F1-score (8) is described as a trade-off metric that is a mixed of recall and precision [51]. IOU (9) refers to the amount of common pixels between the target and prediction masks divided by the entire amount of existing pixels across both masks [21]. The metrics were calculated based on false negative (FN), true negative (TN),

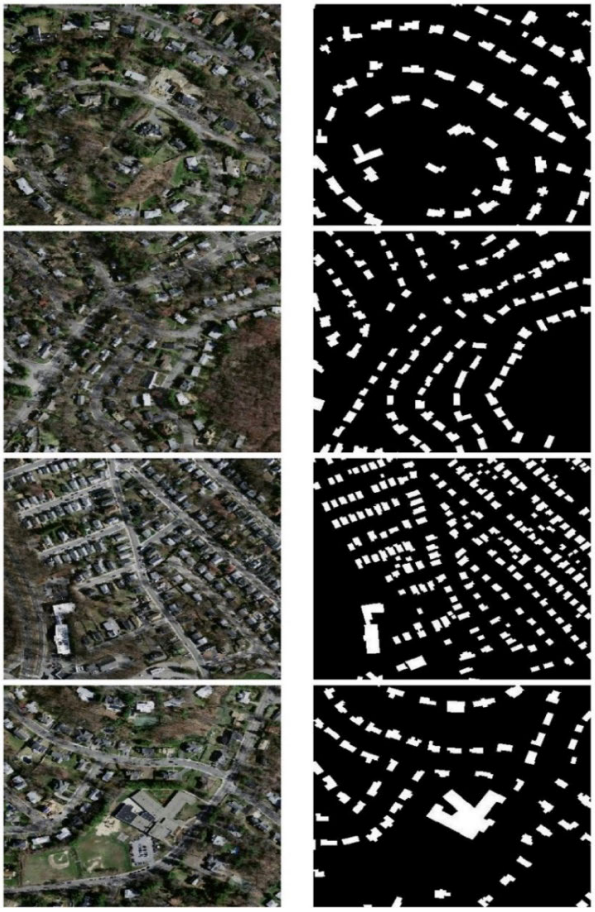

FIGURE 5. Some example images in our building dataset. The original images and corresponding reference maps are shown in the first and second columns, respectively.

false positive (FP), and true positive (TP) as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$
$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad IOU = \frac{TP}{TP + FP + FN}$$

C. COMPARISON OF BUILDING SEGMENTATION ALGORITHMS

We compared the suggested GAN model with other state-of-the-art approaches to verify the performance of the suggested technique in building segmentation. We selected classification-based and CNN-based approaches such as Sameen and Pradhan [52] and SegNet model without BConvLSTM [15] for comparison, as the proposed model is a pixel-wise segmentation method. The proposed SegNet+BConvLSTM model was also used for the generator part in our proposed GAN model. Therefore, the difference between the GAN model (adversarial training) and the SegNet model (non-adversarial training) for building extraction could be observed.

D. QUANTITATIVE AND QUALITATIVE RESULTS

For evaluating the effectiveness of the suggested GAN model for building extraction, we present the visual building extraction results of the proposed model and other approaches in Figure 6. The results show that all the extraction approaches could alleviate the effect of obstacles to a certain degree on the basis of the spatial information consideration. However, the outcomes of the Sameen and Pradhan's approach [52] presented less FNs and more FPs that are illustrated by the yellow and blue colors, respectively. In contrast, a segmentation map with a similar resolution as the input image could be generated using SegNet model and proposed SegNet+BConvLSTM model that uses deconvolutional layers. Therefore, the accuracy of boundary information obtained by these models is higher than that of [52]. The proposed SegNet model with BConvLSTM were used in the generative part of the GAN model for adversarial training. Therefore, the segmentation map obtained by the GAN model is smoother than that of the SegNet model, with fewer FPs.

TABLE 1. Quantitative outcomes achieved by the proposed GAN model and other comparative approaches.

		Sameen and Pradhan[52]	SegNet [15]	SegNet +BConvLSTM	Prop-GAN
Image1	Recall	0.8941	0.9635	0.9647	0.9728
	Precision	0.9307	0.9395	0.9488	0.9569
	F1-score	0.9120	0.9514	0.9567	0.9648
	IOU	0.8382	0.9072	0.9169	0.9319
Image2	Recall	0.8482	0.9332	0.9628	0.9652
	Precision	0.9468	0.9688	0.9500	0.9637
	F1-score	0.8948	0.9507	0.9563	0.9644
	IOU	0.8095	0.9060	0.9163	0.9313
Image3	Recall	0.8977	0.9613	0.9475	0.9679
	Precision	0.9045	0.9149	0.9462	0.9519
	F1-score	0.9011	0.9375	0.9469	0.9598
	IOU	0.8200	0.8824	0.8990	0.9227
Image4	Recall	0.9060	0.9481	0.9670	0.9718
	Precision	0.9512	0.9596	0.9469	0.9663
	F1-score	0.9281	0.9538	0.9568	0.9690
	IOU	0.8658	0.9116	0.9172	0.9398
Image5	Recall	0.9431	0.9401	0.9485	0.9770
	Precision	0.9212	0.9507	0.9488	0.9574
	F1-score	0.9320	0.9454	0.9487	0.9671
	IOU	0.8726	0.8964	0.9023	0.9362
Average	Recall	0.8978	0.9492	0.9581	0.9709
	Precision	0.9308	0.9467	0.9481	0.9592
	F1-score	0.9136	0.9477	0.9530	0.9650
	IOU	0.8412	0.9007	0.9103	0.9323

The quantitative outcomes of the proposed GAN model and other comparison approaches are shown in Table 1 to highlight the effectiveness of the proposed GAN model in building extraction. The segmentation precision of the comparison methods for the five test images is shown in the first four rows of the Table 1 and the average performance is shown in the last row of the Table 1.

Table 1 shows that the accuracy of all defined metrics of the Sameen and Pradhan's method is less than those of the other methods for whole images. This is because the method predicted less FNs and more FPs, especially where the buildings are covered by other obstacles that led to a less accuracy. The SegNet model [15] could achieve higher accuracy than the Sameen and Pradhan's and a non-noisy segmentation map, while the proposed SegNet+BConvLSTM

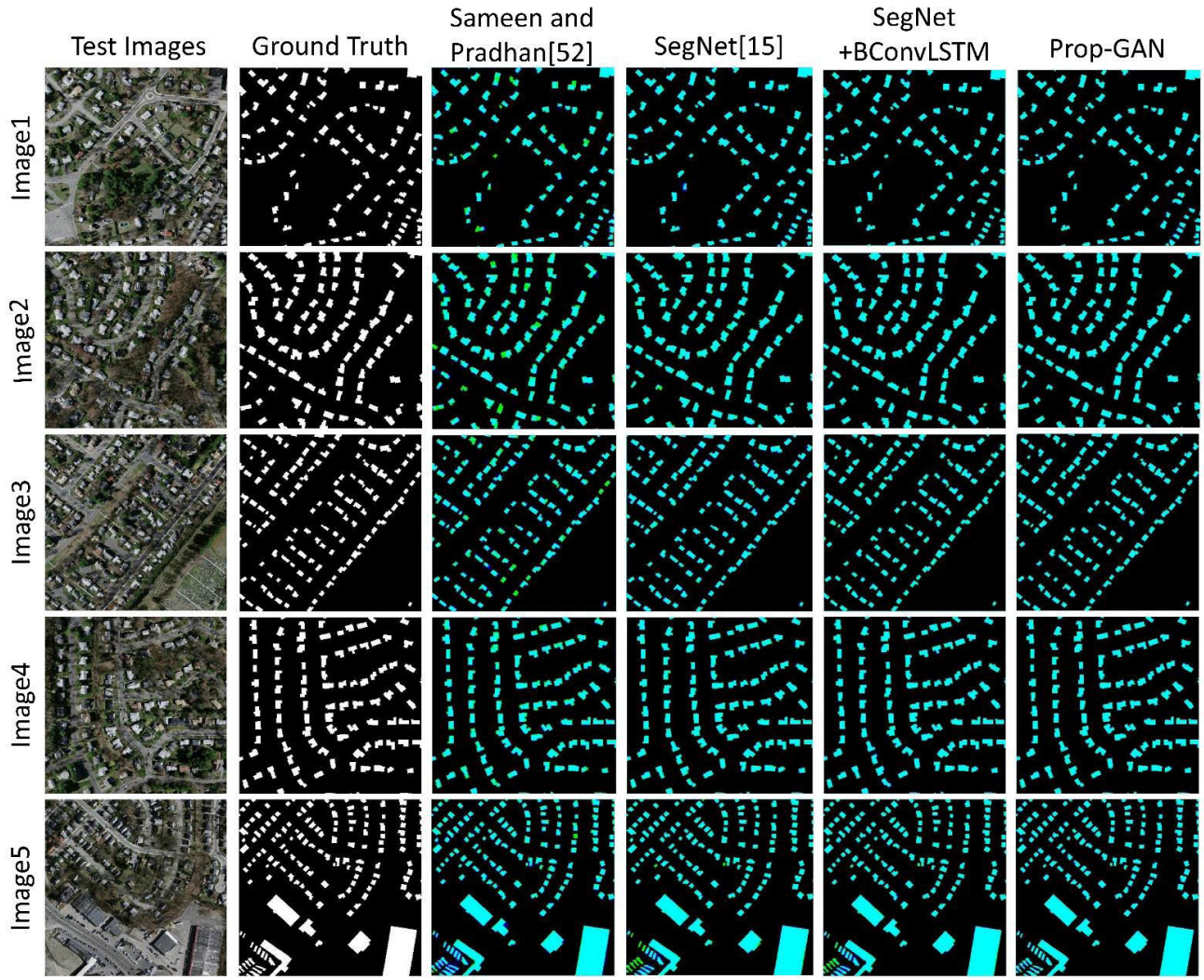

FIGURE 6. Building extraction outcomes achieved by GAN model and other state-of-the-art methods. First and second rows show the original images and corresponding labels. The third, fourth, fifth and sixth rows depict the results achieved by the Sameen and Pradhan's method [52], SegNet [15], proposed SegNet+BConvLSTM, and proposed GAN, respectively. The black (background), white, yellow, and blue colors represent TNs, TPs, FPs, and FNs, respectively.

model even could achieve better results than both Sameen and Pradhan's and SegNet models, making it the second-best method. The proposed GAN model could achieve the highest precision for whole images among the approaches. The average precision obtained by the GAN model for IOU is nearly 9.11%, 3.16% and 2.2% higher than those by the Sameen and Pradhan's method, SegNet method and the proposed SegNet+BConvLSTM model, respectively, indicating that the proposed GAN model (adversarial training) could improve the outcomes and verifying the superiority of the proposed model for building extraction over the other methods.

E. DISCUSSION

We also compared the quantitative and qualitative outcomes attained by the proposed GAN model with more pre-existing state-of-the-art classifiers such as improved fully convolutional model with conditional random fields (FCN-CRFs) [37] and generative adversarial framework with mechanism of spatial and channel attention (GAN-SCA) [53] to prove the superiority of the method in producing high-quality building segmentation map. Note that the outcomes for the other methods were adopted from the main published papers, whereas the suggested GAN model has been performed on an experimental building data. The qualitative results obtained by the proposed GAN network and other deep learning-based models are shown in Table 2. It is observed from the table that the proposed SegNet+BConvLSTM and GAN models could generally obtain high accuracy for F1-score. The table shows that the GAN-SCA [53] could obtain higher F1-score than the proposed SegNet model by a margin of only 0.4% and it is ranked in the second best model for building extraction. However, the proposed GAN model in this study could improve the results to 0.45% and achieve higher F1-score. Moreover, the proposed GAN model could yield higher F1-score than those of [37]. Besides, the qualitative results obtained by the

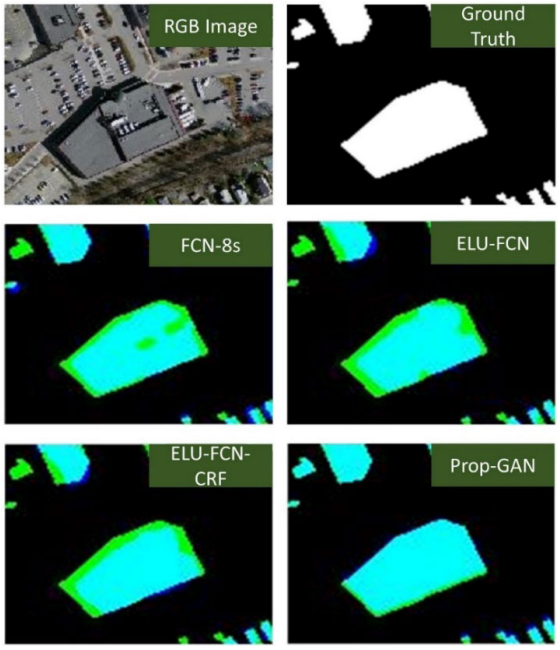

FIGURE 7. Visualization results of the proposed GAN approach and other state-of-the-art pre-existing classifiers. FPs and FNs are illustrated in blue and yellow colors.

TABLE 2. Results obtained by the proposed GAN network and other pre-existing classifiers based on f1-score.

Model	FCN [37]	FCN-ELU [37]	FCN-ELU-CRFs [37]	GAN-SCA [53]	SegNet+ BConvLSTM	Prop-GAN
F1-score	0.9522	0.9533	0.9536	0.9636	0.9596	**0.9681**

comparative algorithms are shown in Figure 7. As it can be seen, the FCN model could not detect the edges of buildings and predicted more FPs in the edges that lead to achieving lower accuracy than the proposed model; and consequently produced a low-quality building segmentation map. In contrast, it is clear that the proposed GAN model could generate a non-noisy building segmentation map with preserving the boundary information compared to other methods that predicted more FP and less FN pixels even under occlusions of trees and cars. This is because the adversarial training method obtained more consistent building segmentation results than other approaches by enforcing long-range spatial label contiguity. Also, by adding BConvLSTM modules to the decoding path of SegNet model that was used for generator part of GAN model mixes the feature maps exploited from the corresponding contracting part and the prior expanding up-sampling layer in a non-linear way which could remove salt-and-pepper noise and improve the performance of the network.

V. CONCLUSION

In the current study, we proposed a novel end-to-end convolutional neural architecture based on the GAN model to extract buildings from high-resolution remote sensing images. We used adversarial training to improve the accuracy of the segmentation map. Additionally, this paper aimed to design an encoder–decoder model called SegNet with BCovLSTM function for the generative part of the proposed GAN model to produce the pixel-wise classification map and improve the accuracy of building extraction. BCovLSTM module was added to the expansive part of the SegNet model to mix encoded features with higher resolution and local information and decoded features with more semantic information, which eliminate the noises and improve the performance of the model in building detection under complicated backgrounds. Another reason that the proposed GAN model could even extract buildings in the complex areas, where the buildings surrounded by trees, cars, shadows, and other occlusions is that the adversarial training technique execute long-range spatial label adjacency to obtain the high-quality results of building detection under complex situations and occlusions. We applied the common measurement factors called precision, recall, F1-score and IOU to show the effectiveness of the proposed method for building extraction, obtaining an average of 96.81% for F1-score, indicating that the suggested model could perform accurate building extraction and non-noisy segmentation map acquisition. In addition, we compared the proposed GAN model with some of the state-of-the-art approaches (the results shown in Figure 6 and 7) to highlight the advantages of the suggested technique in building extraction. The outcomes confirm the superiority of the proposed technique for building extraction over other methods and the capability of the GAN model to achieve the best visualization and quantitative performance.

AUTHOR CONTRIBUTIONS

Conceptualization, A.A. and B.P.; methodology and formal analysis, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, B.P.; supervision, B.P.; funding – B.P. and A.A.A., All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] H. R. R. Bakhtiari, A. Abdollahi, and H. Rezaeian, "Semi automatic road extraction from digital images," *Egyptian J. Remote Sens. Space Sci.*, vol. 20, no. 1, pp. 117–123, Jun. 2017, doi: 10.1016/j.ejrs.2017.03.001.

[2] 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. (IGARSS)*, Jul. 2016, pp. 1591–1594, doi: 10.1109/IGARSS.2016.7729406.

[3] S. Paisitkriangkrai, J. Sherrah, P. Janney, and V.-D. Hengel, "Effective semantic pixel labelling with convolutional networks and conditional random fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 36–43, doi: 10.1109/CVPRW.2015.7301381.

[4] J. Peng and Y. C. Liu, "Model and context-driven building extraction in dense urban aerial images," *Int. J. Remote Sens.*, vol. 26, no. 7, pp. 1289–1307, Apr. 2005, doi: 10.1080/01431160512331326675.

[5] Z. J. Liu, J. Wang, and W. P. Liu, "Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic Hough transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4, Jul. 2005, pp. 2250–2253, doi: 10.1109/IGARSS.2005.1525421.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, 2012, pp. 1097–1105, doi: 10.1145/3065386.

[7] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833, doi: 10.1007/978-3-319-10590-1_53.
[8] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1265–1274, doi: 10.1109/CVPR.2015.7298731.
[9] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1377–1385. [Online]. Available: https://arxiv.org/abs/1509.02634
[10] C.-A. Brust, S. Sickert, M. Simon, E. Rodner, and J. J. A. P. A. Denzler, "Convolutional patch networks with spatial prior for road detection and urban scene understanding," 2015, *arXiv:1502.06344*. [Online]. Available: https://arxiv.org/abs/1502.06344
[11] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062
[12] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1742–1750, doi: 10.1109/ICCV.2015.203.
[13] S. Hong, H. Noh, and B. Han, "Decoupled deep neural network for semi-supervised semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1495–1503. [Online]. Available: https://arxiv.org/abs/1506.04924
[14] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528, doi: 10.1109/ICCV.2015.178.
[15] 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, doi: 10.1109/TPAMI.2016.2644615.
[16] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, "Multi-scale object detection in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 3–22, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.04.003.
[17] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018, doi: 10.1109/TGRS.2017.2783902.
[18] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. D. 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, Aug. 2017, doi: 10.1016/j.isprsjprs.2017.05.002.
[19] B. Pradhan, H. A. H. Al-Najjar, M. I. Sameen, M. R. Mezaal, and A. M. Alamri, "Landslide detection using a saliency feature enhancement technique from LiDAR-derived DEM and orthophotos," *IEEE Access*, vol. 8, pp. 121942–121954, 2020, doi: 10.1109/ACCESS.2020.3006914.
[20] B. Pradhan, H. A. H. Al-Najjar, M. I. Sameen, I. Tsang, and A. M. Alamri, "Unseen land cover classification from high-resolution orthophotos using integration of zero-shot learning and convolutional neural networks," *Remote Sens.*, vol. 12, no. 10, p. 1676, May 2020, doi: 10.3390/rs12101676.
[21] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," *IEEE Access*, vol. 8, pp. 179424–179436, 2020, doi: 10.1109/ACCESS.2020.3026658.
[22] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.
[23] 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, p. 1444, May 2020, doi: 10.3390/rs12091444.
[24] Q. Jiang, L. Cao, M. Cheng, C. Wang, and J. Li, "Deep neural networks-based vehicle detection in satellite images," in *Proc. Int. Symp. Bioelectronics Bioinf. (ISBB)*, Oct. 2015, pp. 184–187, doi: 10.1109/ISBB.2015.7344954.
[25] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," 2016, *arXiv:1611.08408*. [Online]. Available: http://arxiv.org/abs/1611.08408
[26] Q. Shi, X. Liu, and X. Li, "Road detection from remote sensing images by generative adversarial networks," *IEEE Access*, vol. 6, pp. 25486–25494, 2017, doi: 10.1109/ACCESS.2017.2773142.
[27] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, Graduate Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.
[28] Y. Shu, "Deep convolutional neural networks for object extraction from high spatial resolution remotely sensed imagery," Ph.D. dissertation, Dept. Geography, Univ. Waterloo, Waterloo, ON, Canada, 2014.
[29] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," in *Proc. 8th Image Mach. Vis. Appl.*, vol. 9405, Feb. 2015, p. 94050, doi: 10.1117/12.2083273.
[30] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks," *Electron. Imag.*, vol. 2016, no. 10, pp. 1–9, Feb. 2016, doi: 10.2352/issn.2470-1173.2016.10.robvis-392.
[31] Z. Huang, G. Cheng, H. Wang, H. Li, L. Shi, and C. Pan, "Building extraction from multi-source remote sensing images via deep deconvolution neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1835–1838, doi: 10.1109/IGARSS.2016.7729471.
[32] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017, doi: 10.1109/TGRS.2016.2612821.
[33] J. Yuan and A. M. Cheriyadat, "Learning to count buildings in diverse aerial scenes," in *Proc. 22nd ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL)*, 2014, pp. 271–280, doi: 10.1145/2666310.2666389.
[34] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 1873–1876, doi: 10.1109/IGARSS.2015.7326158.
[35] K. Bittner, S. Cui, and P. Reinartz, "Building extraction from remote sensing data using fully convolutional networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.-ISPRS Arch.*, vol. 42, no. W1, pp. 481–486, 2017, doi: 10.5194/isprs-archives-XLII-1-W1-481-2017.
[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[37] S. Shrestha and L. Vanneschi, "Improved fully convolutional network with conditional random fields for building extraction," *Remote Sens.*, vol. 10, no. 7, p. 1135, Jul. 2018, doi: 10.3390/rs10071135.
[38] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, vol. 39, no. 4, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.
[39] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[41] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building extraction in very high resolution remote sensing imagery using deep learning and guided filters," *Remote Sens.*, vol. 10, no. 1, p. 144, Jan. 2018, doi: 10.3390/rs10010144.
[42] H. T. Aung, S. H. Pha, and W. Takeuchi, "Building footprint extraction in yangon city from monocular optical satellite image using deep learning," *Geocarto Int.*, pp. 1–21, Mar. 2020, doi: 10.1080/10106049.2020.1740949.
[43] Z. Shao, P. Tang, Z. Wang, N. Saleem, S. Yam, and C. Sommai, "BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images," *Remote Sens.*, vol. 12, no. 6, p. 1050, Mar. 2020, doi: 10.3390/rs12061050.
[44] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019, doi: 10.1109/TGRS.2019.2902431.
[45] X. Li, X. Yao, and Y. Fang, "Building-A-nets: Robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3680–3687, Oct. 2018, doi: 10.1109/JSTARS.2018.2865187.
[46] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Trans. Geosci. Remote Sens.*, early access, Jun. 18, 2020, doi: 10.1109/TGRS.2020.3000296.

[47] L. Ding, M. H. Bawany, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "A novel deep learning pipeline for retinal vessel detection in fluorescein angiography," *IEEE Trans. Image Process.*, vol. 29, pp. 6561–6573, 2020, doi: 10.1109/TIP.2020.2991530.
[48] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," 2018, *arXiv:1801.02143*. [Online]. Available: http://arxiv.org/abs/1801.02143
[49] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam, "Pyramid dilated deeper convlstm for video salient object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 715–731, doi: 10.1007/978-3-030-01252-6_44.
[50] N. Ghasemkhani, S. S. Vayghan, A. Abdollahi, B. Pradhan, and A. Alamri, "Urban development modeling using integrated fuzzy systems, ordered weighted averaging (OWA), and geospatial techniques," *Sustainability*, vol. 12, no. 3, p. 809, Jan. 2020, doi: 10.3390/su12030809.
[51] S. Wang, X. Hou, and X. Zhao, "Automatic building extraction from high-resolution aerial imagery via fully convolutional encoder-decoder network with non-local block," *IEEE Access*, vol. 8, pp. 7313–7322, 2020, doi: 10.1109/ACCESS.2020.2964043.
[52] M. I. Sameen and B. Pradhan, "A novel road segmentation technique from orthophotos using deep convolutional autoencoders," *Korean J. Remote Sens*, vol. 33, no. 4, pp. 423–436, 2017, doi: 10.7780/kjrs.2017.33.4.8.
[53] X. Pan, F. Yang, L. Gao, Z. Chen, B. Zhang, H. Fan, and J. Ren, "Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms," *Remote Sens.*, vol. 11, no. 8, p. 917, Apr. 2019, doi: 10.3390/rs11080917.

ABOLFAZL ABDOLLAHI received the B.Sc. degree from the Ferdowsi University of Mashhad, Iran, and the M.Sc. degree in GIS and remote sensing from the Kharazmi University of Tehran, Iran. He is currently pursuing the Ph.D. degree with the Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), University of Technology Sydney (UTS). His research interests contain the application of advanced machine learning approaches and deep learning-based networks for remote sensing image classification, image segmentation, feature extraction, and GIS maps database updating. He has published numerous peer-reviewed papers on the application of machine learning approaches. He was rewarded the International Research Scholarship and UTS Presidents' Scholarship for the current course in 2018.

BISWAJEET PRADHAN (Senior Member, IEEE) received the Habilitation degree in remote sensing from the Dresden University of Technology, Germany, in 2011. He is currently the Director of the Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT. He is also the Distinguished Professor with the University of Technology Sydney. He is also an internationally established Scientist in the fields of geospatial information systems (GIS), remote sensing and image processing, complex modeling/geo-computing, machine learning and soft-computing applications, natural hazards, and environmental modeling. He has widely travelled abroad, visiting more than 52 countries to present his research findings. Since 2015, he has been serving as the Ambassador Scientist for the Alexander Humboldt Foundation, Germany. He has published over 550 articles of which more than 475 have been published in science citation index (SCI/SCIE) technical journals. He has authored eight books and 13 book chapters. He was a recipient of the Alexander von Humboldt Fellowship from Germany. He received 55 awards in recognition of his excellence in teaching, service, and research, since 2006. He was also a recipient of the Alexander von Humboldt Research Fellowship from Germany. From 2016 to 2018, he was listed as the World's Most Highly Cited Researcher by Clarivate Analytics Report as one of the world's most influential mind. In 2018, 2019, and 2020, he was awarded as the World Class Professor by the Ministry of Research, Technology and Higher Education, Indonesia. He is also an Associate Editor and an Editorial Member of more than eight ISI journals.

SHILPA GITE received the M.Tech. (IT) degree from Bharati Vidyapeeth, Pune, in 2013, and the Ph.D. degree in deep learning for assistive driving from Symbiosis International (Deemed University), Pune, India, in 2019. She is currently working as an Assistant Professor with the Computer Science Department, Symbiosis Institute of Technology, Pune. She has around 12 years of teaching experience. Her research areas include deep learning, machine learning, and computer vision. She is currently guiding Ph.D. students in biomedical imaging, self-driving cars, and natural language processing areas. She has published more than 40 research papers in International journals and 15 Scopus indexed International Conferences. Out of her 40 Journals, 12 papers are Scopus/SCI indexed journals which include 1 Q1 and 1 Q2 journal. She is also a recipient of the Best Paper Award at the 11th IEMERA Conference held virtually at Imperial College, London, in October 2020.

ABDULLAH ALAMRI received the B.S. degree in geology from King Saud University, in 1981, the M.Sc. degree in applied geophysics from the University of South Florida, Tampa, FL, USA, in 1985, and the Ph.D. degree in earthquake seismology from the University of Minnesota, Minneapolis, MN, USA, in 1990. He is currently a Professor of Earthquake Seismology, and the Director of the Seismic Studies Center, King Saud University (KSU). He is the President of the Saudi Society of Geosciences and the Editor-in-Chief of the *Arabian Journal of Geosciences* (AJGS). He is a member of the Seismological Society of America, American Geophysical Union, European Association for Environmental and Engineering Geophysics, Earthquakes Mitigation in the Eastern Mediterranean Region, National Commission for Assessment and Mitigation of Earthquake Hazards in Saudi Arabia, and Mitigation of Natural Hazards Com at Civil Defense. His research interests are in the area of crustal structures and seismic micro zoning of the Arabian Peninsula. His recent projects also involve applications of EM and MT in deep groundwater exploration of Empty Quarter and geothermal prospecting of volcanic Harrats in the Arabian shield. He has published more than 150 research papers, achieved more than 45 research projects, as well as authored several books and technical reports. He is a principal and co-investigator in several national and international projects (KSU, KACST, NPST, IRIS, CTBTO, US Air force, NSF, UCSD, LLNL, OSU, PSU, and Max Planck). He has also chaired and co-chaired several SSG, GSF, RELEMR workshops, and forums in the Middle East. He obtained several worldwide prizes and awards for his scientific excellence and innovation.