Journal of World Architecture, 2024, Volume 8, Issue 2
https://ojs.bbwpublisher.com/index.php/JWA
Online ISSN: 2208-3499
Print ISSN: 2208-3480

Research on the Preservation Method of Traditional Village Roof Information: A Case Study of Gubeikou Village

Mengchen Wang*

College of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Corresponding author: Mengchen Wang, 201604010412@stu.bucea.edu.cn

Abstract: Traditional Chinese villages serve as crucial repositories of traditional culture. However, In China, the urgent task of preserving information about traditional village architecture has arisen due to the degradation of these villages' appearance caused by rapid urbanization in recent years. This paper proposes a method for preserving information about traditional village rooftops based on high spatial resolution remote sensing imagery. Leveraging an improved Mask R-CNN model, the method conducts target recognition on the rooftops of traditional village buildings and generates vectorized representations of these rooftops. The precision rate, recall rate, and F1-score achieved in the experimental results are 93.26%, 86.33%, and 92.02%, respectively. These findings indicate the effectiveness of the proposed method in preserving information about traditional village architecture and providing a viable approach to support the sustainable development of traditional villages in China.

Keywords: Traditional villages; Building rooftops; HSRRS; Mask R-CNN; Instance segmentation

Online publication: May 21, 2024

1. Introduction

With the rapid development of the economy, traditional villages in China have been subjected to urbanization, changes in economic models, and weak awareness of historical and cultural preservation, resulting in the destruction of traditional buildings. Guiding opinions were issued by the Beijing Municipal People's Government in 2018, outlining key tasks for strengthening the protection and development of traditional villages, with a focus on enhancing the preservation of traditional architecture, including a comprehensive investigation of the distribution of traditional buildings within traditional villages, and emphasizing the importance of preserving information about traditional village buildings.

High-resolution remote sensing imagery, characterized by rich details of ground objects and easy data acquisition, can effectively meet the basic data requirements for preserving information about rooftops in traditional Chinese villages on a large scale. Currently, traditional rooftop identification techniques based on high-resolution remote sensing imagery primarily rely on deep learning neural network models, with scholars both domestically and internationally achieving a series of results [1-7]. Tejeswari *et al.* [6] generated training data for specific regional locations using the Google API and performed high-precision extraction of urban buildings based on the Mask R-CNN model. Wang [7] combined the Path Aggregation Feature Pyramid Network (PAFPN) and Atlas Spatial Pyramid Pooling (ASPP) with the Mask R-CNN model and constructed eight sample sets of different color scales for traditional villages in Beijing. However, due to the limitations of convolutional models, errors and omissions in target recognition still exist, underscoring the importance of seeking a simple and efficient method for preserving information about traditional village rooftops. This paper proposes a method for preserving information about traditional village building rooftops based on an improved Mask R-CNN model[8], aiming to contribute to the sustainable protection and development of traditional villages.

2. Methodology for traditional village building rooftop recognition

2.1. Methodology for preserving traditional village building rooftop information

Figure 1 illustrates the technical process of the method for preserving information about traditional village building rooftops, which primarily comprises three pivotal stages. Firstly, preprocessing of high-resolution remote sensing image data from the experimental area is conducted, encompassing image enhancement of the cropped tile images; secondly, inputting the experimental data into the pre-trained improved Mask R-CNN model, and concatenating the recognized building rooftop mask data obtained; finally, vectorization processing of the obtained building rooftop grid mask data from the object recognition, including steps such as binarization, rectification, and reclassification, to obtain high-precision vector data of building rooftops.

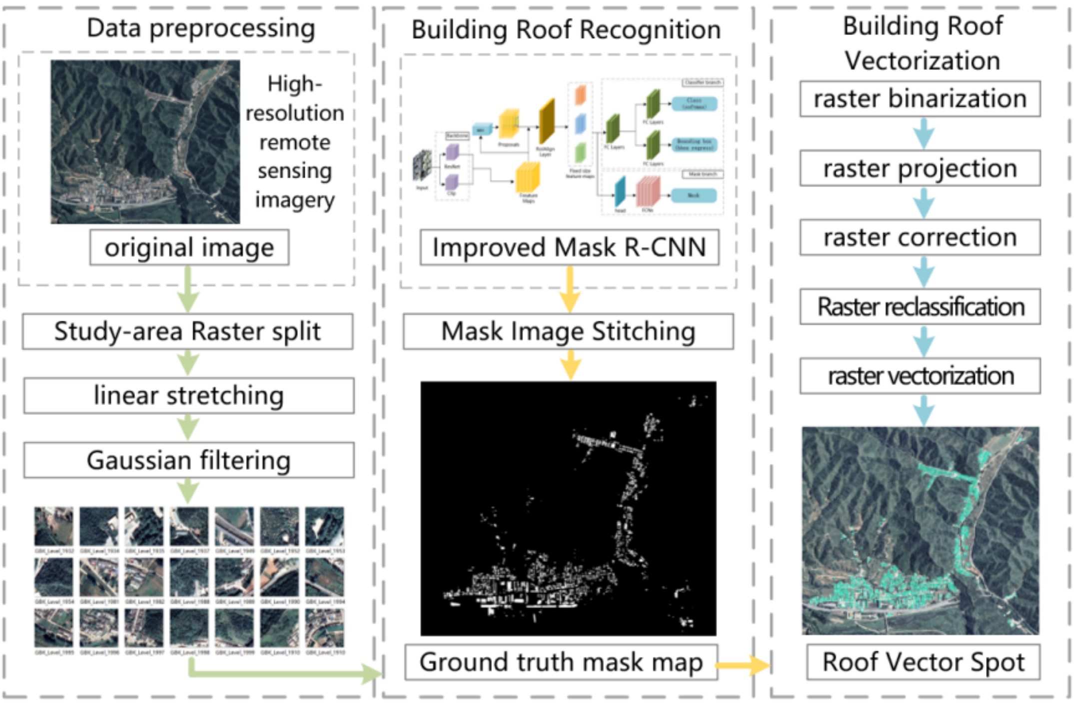

Figure 1. Traditional village building rooftop information preservation methodology technical workflow

2.2. Improved Mask R-CNN model

This study addresses the demand for preserving information regarding the rooftops of traditional village buildings in China. We focus on the misidentification and omission issues of building rooftop targets. We propose an improved Mask R-CNN model [7], which enhances the feature extraction of the Mask R-CNN model [10] using the CLIP [9] model. Additionally, we construct a systematic approach for preserving traditional village building rooftop information by combining the feature analysis of traditional village building rooftops in China and constructing a lineage-based sample set.

As depicted in **Figure 2**, the enhanced Mask R-CNN network model primarily employs the Mask R-CNN instance segmentation network model as its foundational architecture. It integrates the instance segmentation capability of Mask R-CNN [7] by combining the multi-scale representation of the Feature Pyramid Network with the multi-modal interpretability of the CLIP model to construct a semantically enhanced deep convolutional neural network for improving the recognition accuracy of building rooftops. The overall architecture of the improved Mask R-CNN network model is divided into three parts. The first part improved the backbone part, integrating CLIP's visual encoder and text encoder to achieve deep feature extraction and semantic understanding. The second part involves the RPN, responsible for generating region proposals based on the previously extracted feature maps and refining candidate boxes for regions of interest. The generated target candidate boxes are inputted into the RoIAlign layer, which aligns the features obtained from each ROI with the original image ROI region more effectively and converts them into a fixed dimension for subsequent fully connected layers. The third segment comprises three parallel branches: one for classification, another for bounding box regression, and the third for mask generation.

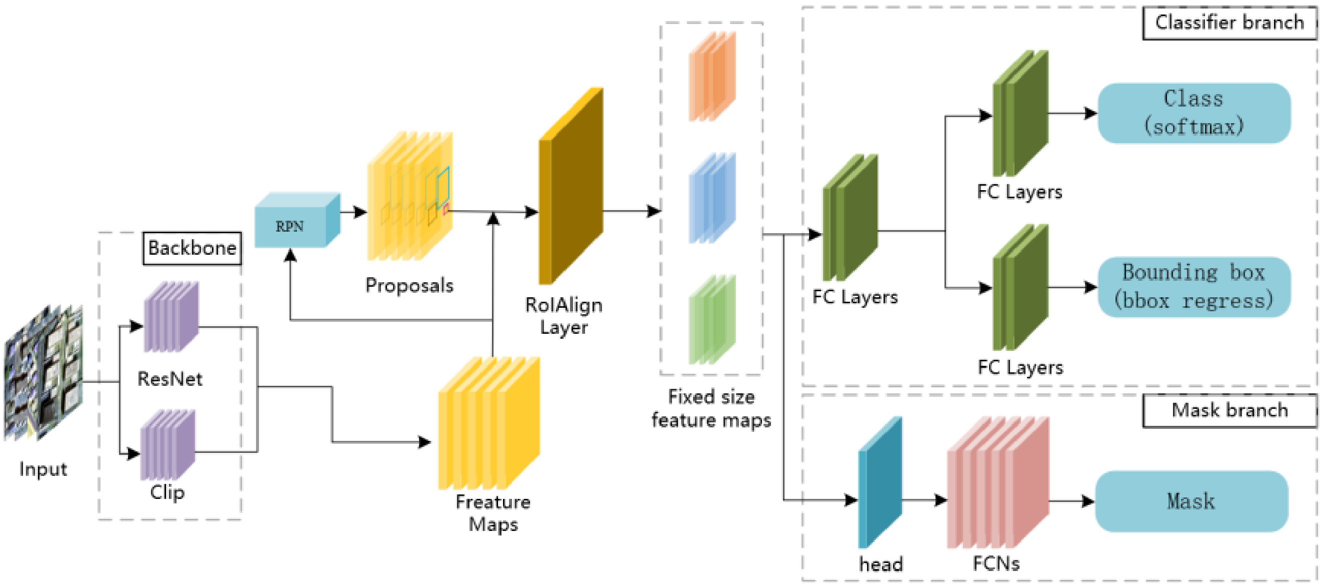

Figure 2. General architecture of Improved Mask R-CNN network model [8]

3. Experimentation and validation

3.1. Experimental area data

To validate the practicality of the high-resolution remote sensing method for extracting traditional building rooftops, this study selected Guobeikou Village in Guobeikou Town, Miyun District, Beijing, as the representative traditional village in Beijing. The data used were 20-level Google Earth imagery, obtained from non-single-source data, with a spatial resolution of 0.27 meters, including channels for the R, G, and B bands. The dataset size was approximately 1.0 GB, covering an area of about 4.36 km². Guobeikou Village is designated as a standardized folk tourism village in Beijing, characterized by a blend of traditional Chinese architecture, tourist-cultural imitation antique buildings, and modern industrial structures. The rooftops display a variety of shapes and textures, as well as significant spectral and geometric variations, rendering them ideal for assessing the viability of the traditional village building rooftop preservation method through high-resolution remote sensing imagery.

3.2. Experiment

3.2.1. Image preprocessing

This study conducts a series of data preprocessing steps on the high-resolution remote sensing data. Firstly, the image within the test area undergoes a split cropping process, resizing it to 512*512 to suit the input requirements of the neural network model for target recognition. Secondly, image enhancement is conducted based on the grayscale histogram, converting the 16-bit image to an 8-bit image through linear stretching. This reduces data volume, enhances processing efficiency, and improves pixel contrast, making roof target features more discernible. Lastly, spatial domain filtering enhances the image, with Gaussian filtering employed to smooth Gaussian noise effects and refine spectral features of building rooftops, thereby enhancing spectral feature heterogeneity in the image.

3.2.2. Building roof recognition

In the building roof recognition experiment conducted in Gubei Village, the operating system utilized is Ubuntu 18.04, and the GPU version employed is GeForce RTX 3090. The high-resolution remote sensing image data of Gubei Village is processed by binning and cropping into 340 binned images to facilitate object-oriented target recognition of building rooftops. Initially, the roofs are categorized into traditional, antique, and modern buildings based on the characteristics of building roofs in traditional Chinese villages. Subsequently, a tailored training dataset is constructed to train the improved Mask R-CNN model [8]. Following this, optimal hyperparameters and network structures of the model are determined through analysis of the validation set. In this paper, roof recognition is performed based on the trained and tuned model [8] oriented to the roofs of buildings in traditional Chinese villages, as shown in part (b) of **Figure 3**.

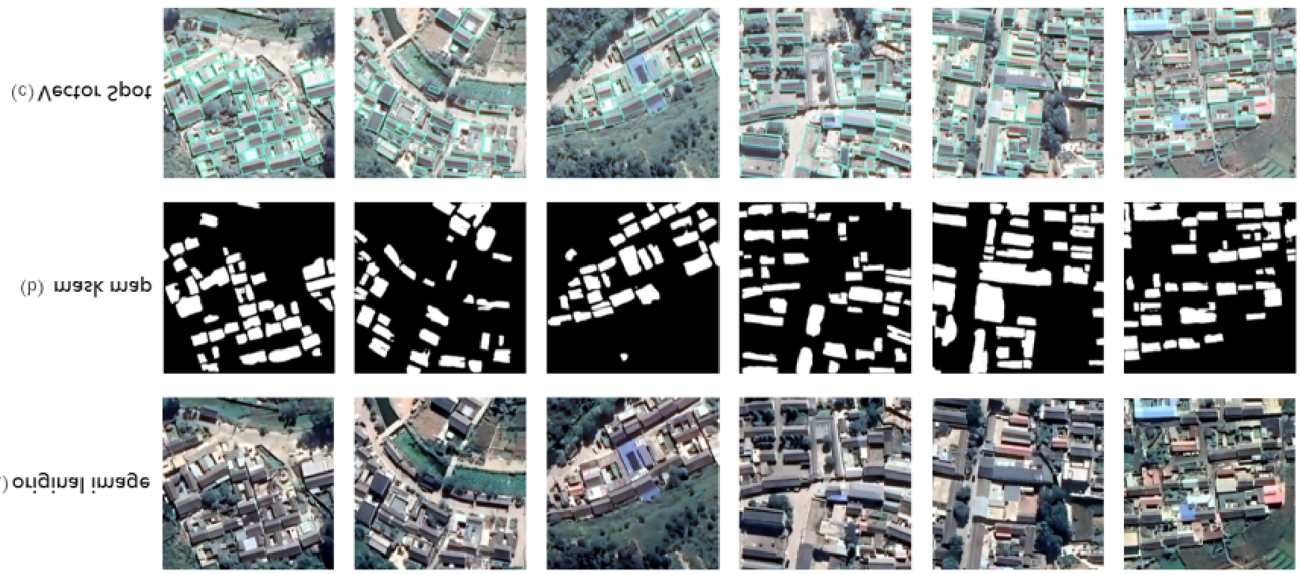

Figure 3. Experimental area target identification results

3.2.3. Building roof vectorization

The process of vectorizing building roofs involves several steps: piecing together small image segments (raster splicing), turning these images into black-and-white to distinguish targets from their background (raster binarization), adjusting these images to fit specific projections and alignments (raster correction), sorting these images into categories (raster reclassification), and finally transforming them into vector forms (raster vectorization). Initially, to boost pre-processing efficiency, images are segmented into smaller pieces, which are then merged back together using splicing algorithms, as depicted in **Figure 3(b)**. Binarization helps isolate the roof areas identified by the deep learning model from their surroundings, simplifying their conversion into vector formats. For raster correction, assigning correct projection and alignment is crucial, especially since the network's output lacks coordinate data, which is vital for subsequent analysis. Reclassification then simplifies the data, facilitating roof extraction, leading to the final step where these prepared images are converted into vector patches. This vectorization is semi-automatic, requiring manual checks to ensure data integrity and accuracy, as indicated by **Figure 3(c)**. This manual validation helps refine the outcomes, ensuring they align closely with real-world structures.

3.2.4. Experimental results and analysis

In this research, we apply three metrics to assess model accuracy specific to target recognition: Precision, Recall, and F1-score, in relation to binary classification outcomes characterized as True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). These metrics are formulated as follows:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{3}$$

As shown in **Figure 4**, the overall image of the Gubei Village test area and the localized method details of its roof vector patches are demonstrated. Gubei Village is located in the mountainous area of Pinggu District, Beijing, with a large amount of vegetation, and the building types include large-scale industrial modern buildings and small- and medium-scale residential houses, which are distributed haphazardly and with diverse morphologies. From the recognition results, it can be seen that the roof-scale symbiosis of buildings in this test area is better while taking into account the recognition of large- and small-scale roofs, which is more compatible with the actual roof morphology. The improved Mask R-CNN model achieved a precision of 0.9850 and a recall of 0.8633 in the sample area, which indicates that the recognition of building roofs in the area is more accurate, with not many false detections, while the lower recall indicates that there are more target miss-detections of the western phenomenon. From the results, it can be seen that the shadow, tree occlusion, and "foreign objects with the same spectrum" led to the phenomenon of small target missed detection, but the recognition results can still detect most of the small targets, and the overall F1-score of 0.9202 is achieved.

After completing the identification of building roofs in Gubei Village, this study extracts the building roof masks, transforms them into vector patches, and supplements them with manual correction means to obtain the thematic vector map of building roofs in traditional villages in Beijing's Gubei Village, as shown in **Figure 4**. Finally, 1,624 building roofs with a total area of 289,885.4484 m^2 were effectively extracted from Gubei Village in Pinggu District, Beijing.

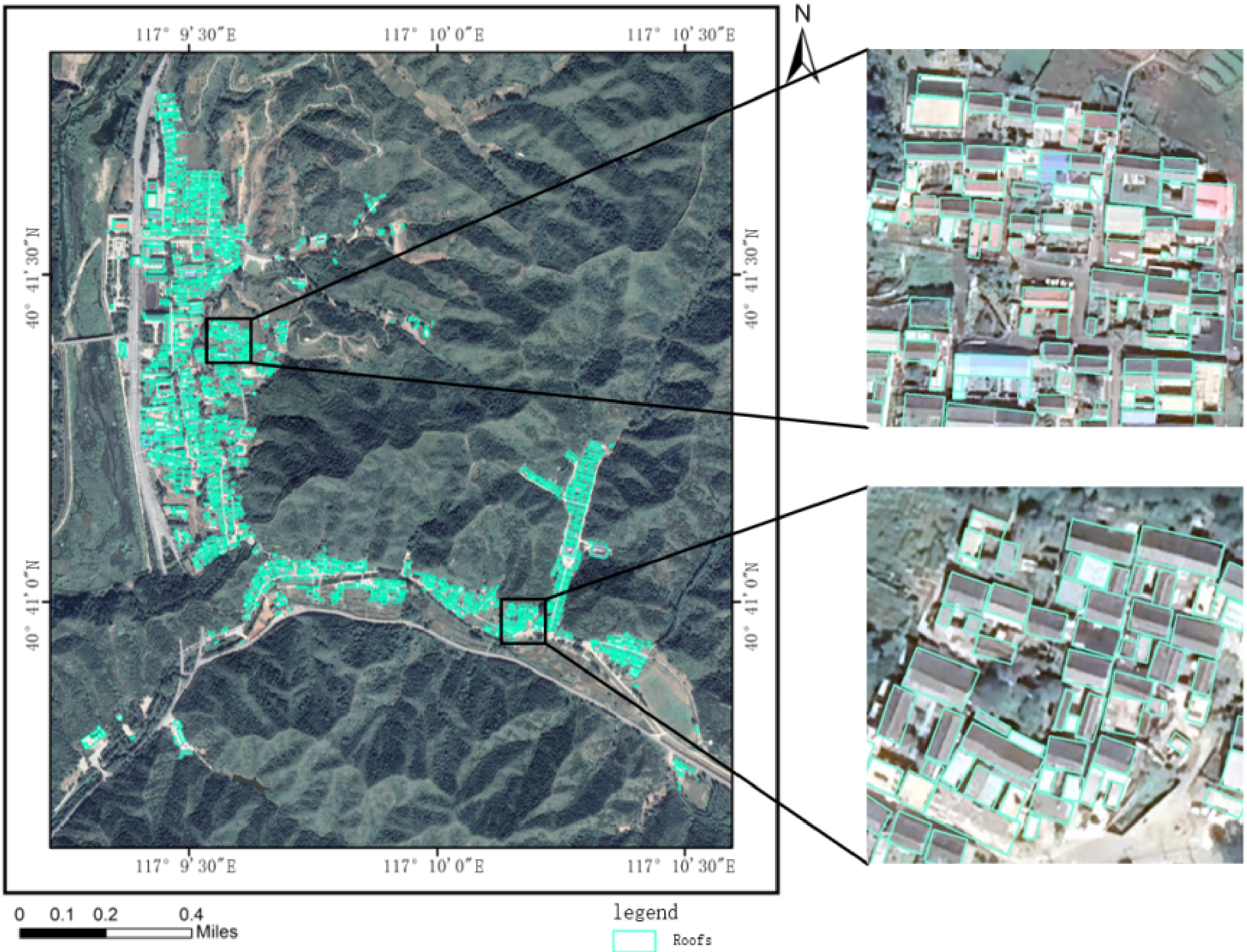

Figure 4. Vector map of building rooftop in Beijing Gubei village

4. Conclusion

In this study, a traditional village building roof information retention method is proposed based on the improved Mask R-CNN model, combined with high-resolution remote sensing images and deep learning techniques. Through experimental validation in Gubeikou Village, Beijing, it is demonstrated that the method has high accuracy and feasibility in traditional village building roof recognition. The experimental results show that the improved Mask R-CNN model exhibits good performance in recognizing building roofs, and at the same time, the building roof vectorization processing method effectively transforms the recognition results into vector patches that can be further analyzed and applied. This study provides a simple and fast method for the protection and information retention of traditional villages in China, which is expected to provide important support for the sustainable development of traditional villages. In the future, the algorithm can be further optimized to improve the recognition accuracy and combine with more geographic information data to expand the application of the method in different regions.

Disclosure statement

The author declares no conflict of interest.

References

[1] Xiong Y, Chen Q, Zhu M, et al., 2020, Accurate Detection of Historical Buildings Using Aerial Photographs and Deep Transfer Learning, IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), 1592–1595.

[2] Liu J, Wang S, Hou X, et al., 2020, A Deep Residual Learning Serial Segmentation Network for Extracting Buildings From Remote Sensing Imagery. International Journal of Remote Sensing, 41(14): 5573–5587.

[3] He K, Gkioxari G, Dollár P, et al., 2017, Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2961–2969.

[4] Li Y, Xu W, Chen H, et al., 2021, A Novel Framework Based on Mask R-CNN and Histogram Thresholding for Scalable Segmentation of New and Old Rural Buildings. Remote Sensing, 13(6): 1070.

[5] Zhan Y, Liu W, Maruyama Y, 2022, Damaged Building Extraction Using Modified Mask R-CNN Model Using Post-Event Aerial Images of the 2016 Kumamoto Earthquake. Remote Sensing, 14(4): 1002.

[6] Tejeswari B, Sharma SK, Kumar M, et al., 2022, Building Footprint Extraction from Space-Borne Imagery using Deep Neural Networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43: 641–647.

[7] Wang W, Shi Y, Zhang J, et al., 2023, Traditional Village Building Extraction Based on Improved Mask R-CNN: A Case Study of Beijing, China. Remote Sensing, 15(10): 2616.

[8] Wang M, Shen L, 2024, High-Resolution Remote Sensing Imagery for the Recognition of Traditional Villages. Journal of Architectural Research and Development, 8(1): 75–83.

[9] Radford A, Kim J W, Hallacy C, et al., 2021, Learning Transferable Visual Models from Natural Language Supervision. International Conference on Machine Learning. PMLR, 8748–8763.

[10] Zhan Y, Liu W, Maruyama Y, 2022, Damaged Building Extraction Using Modified Mask R-CNN Model Using Post-Event Aerial Images of the 2016 Kumamoto Earthquake. Remote Sensing, 14(4): 1002.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.