Data-driven Urbanism: Image Processing Techniques for Urban Analytics.

AL-OBAIDI, Karam <http://orcid.org/0000-0002-4379-6964>, WANG, Jing <http://orcid.org/0000-0002-5418-0217> and HOSSAIN, Mohataz <http://orcid.org/0000-0002-1885-8692>

Published version

AL-OBAIDI, Karam, WANG, Jing and HOSSAIN, Mohataz (2024). Data-driven Urbanism: Image Processing Techniques for Urban Analytics. In: HE, Bao-Jie, PRASAD, Deo, YAN, Li, CHESHMEHZANGI, Ali and PIGNATTA, Gloria, (eds.) International Conference on Urban Climate, Sustainability and Urban Design. Lecture Notes in Civil Engineering (559). Singapore, Springer, 709-720. [Book Section]

Data-Driven Urbanism: Image Processing Techniques for Urban Analytics

Karam M. Al-Obaidi[1](✉), Jing Wang[2], and Mohataz Hossain[1]

[1] Department of the Natural and Built Environment, College of Social Sciences and Arts, Sheffield Hallam University, Sheffield S1 1WB, UK
k.al-obaidi@shu.ac.uk

[2] Department of Computing, College of Business, Technology and Engineering, Sheffield Hallam University, Sheffield S1 1WB, UK

Abstract. Geographic databases provided by open and public sources lack a high degree of accuracy. Although these sources were developed by collecting data from surveys, tracing from aerial imagery and freely licensed geodata sources, their reliability is questionable in testing new concepts for urban analytics and developing solutions for City Information Modelling (CIM). This study aims to examine a method using digital image processing to deliver precise information and accurate data for urban analytics. The research applied algorithmic solutions using content-based image segmentation, which accurately segments roof regions of buildings from aerial images. The study utilised an open access dataset annotated using 72 images grouped into 6 larger tiles from a joint project between Humans in the Loop with the Mohammed Bin Rashid Space Centre in Dubai, the UAE. The results show the efficiency of extracting buildings and their detailed features in an urban context. Finally, the study demonstrates the reliability of using the Base UNet model and the ResNet-based UNet, in analyzing urban aerial images.

Keywords: Urban Analytics · Image Processing · Segmentation · Geodata · Aerial Imagery · Buildings · Roofs

1 Introduction

Climate change and global warming are projected to intensify environmental risks in cities where new approaches to support climate resilience are needed [1]. Urban analytics is the practice of understanding urban and city processes and contexts, currently emerging as an interdisciplinary field, which is gaining popularity in exploring new advances in using Geodata and computational methods [2, 3]. Geographic databases have been widely used and practised in built environment studies to examine and assess urban scenarios [4]. These databases have played a role in defining new challenges in urban studies, such as population, infrastructures, environments, and policies [5].

The development of information technology has offered new possibilities to introduce advanced solutions for urban and city contexts. City Information Modelling (CIM) is a new movement that extends from the development of Building Information Modelling (BIM), a well-established process to provide smart visual models. CIM uses spatial

B.-J. He et al. (Eds.): UCSUD 2023, LNCE 559, pp. 709–720, 2025.
https://doi.org/10.1007/978-981-97-8401-1_51

data from remote sensing and a geographic information system (GIS) by exchanging and sharing information associated with images and textures based on different levels of detail [6].

GIS and remote sensing practices demonstrated limitations in examining specific physical parameters in urban areas [7]. These include the accuracy of identifying, classifying, and quantifying physical characteristics, such as boundaries, materials, textures, and colours. In addition, the reliability of open geographic databases is questionable in testing for urban analytics and developing optimum solutions for CIM.

The integration of computer vision and image processing technologies provides a promising avenue to enhance the capabilities of CIM in gathering valuable data for urban analysis and decision-making. One possible solution is to use Artificial Intelligence (AI) to identify qualifiable information techniques such as image segmentation. Such an approach helps to examine aerial and satellite images and obtain relevant information. By applying these algorithms, it becomes feasible to identify and categorise physical elements such as roof areas and boundaries of buildings. Employing AI helps to address issues related to Urban Heat Island (UHI), energy consumption, and carbon emissions in urban contexts [8]. As a result, this research aims to explore an approach using digital image processing to accurately segment roof regions in aerial images to generate reliable information and accurate data for urban analytics that could help to develop efficient solutions for CIM.

2 Literature Review

Content-based image segmentation involves partitioning an image into meaningful regions based on the content or visual characteristics of the pixels. It has various applications in architecture and engineering, particularly in aerial and satellite image processing. It has the potential to extract buildings, roads, and other infrastructure elements from images, enabling detailed analysis and monitoring of urban areas. Studies in this area have focused on enhancing the accuracy of content-based image segmentation algorithms. Machine learning techniques, such as deep neural networks, have been used to enhance the segmentation results [9]. Additionally, the integration of other data sources, such as LiDAR or GIS data, has been explored to improve the accuracy of segmentation and information about the urban environment [10].

In the last decade, AI and machine learning techniques have revolutionised the area of satellite image analysis. The integration of deep learning algorithms, incorporating convolutional neural networks (CNNs) [11, 12] and recurrent neural networks (RNNs) [13], has brought about significant advancements in the accuracy and efficiency of analysing satellite imagery. These algorithms autonomously acquire knowledge by identifying patterns, features, and objects within satellite images. As a result, deep learning in satellite image analysis has effectively enhanced tasks such as classification, segmentation, and detection, leading to more precise and reliable results. However, this application has limitations when accurately segmenting complex and irregular shapes, such as building edges and road networks from satellite imagery and aerial photography. To overcome these limitations, researchers have turned to deep learning architectures like UNet [14].

In satellite image processing, UNet has shown promising results in segmenting buildings, roads, and other urban features [15]. By training the network on a large dataset of labelled satellite images, it can learn to identify and segment different objects of interest with high accuracy. However, training UNet requires a substantial amount of labelled data and computational resources. The network needs to be trained on a diverse set of satellite images to generalise well to different urban environments. Additionally, hyper-parameter tuning and data augmentation techniques may be necessary to optimise the segmentation performance. In addition, UNet segmentation has limitations when dealing with details such as the roofs of each building and smaller image objects. Due to its architecture and receptive field size, UNet may struggle to accurately segment these fine-grained elements. Additional techniques, such as post-processing or incorporating higher-resolution imagery, may be required to address this challenge.

3 Methodology

The study applied a quantitative approach to collect and analyse numerical data using different neural network frameworks and semantic segmentation of aerial imagery. In this research, a novel and innovative image segmentation algorithm called "Local Object Enhanced-UNet" has been developed. The main objective of this algorithm is to extract the regions of interest from individual building roofs. The Local Object Enhanced-UNet algorithm takes full advantage of the advanced convolution structure of the UNet while incorporating additional post-processing techniques based on low-level visual object features such as colours, textures, and edges. By leveraging these features, the algorithm achieves remarkable accuracy in identifying detailed roof regions. This breakthrough in image segmentation technology opens new possibilities for precise roof material identification and analysis.

Figure 1 illustrates the system pipeline of the proposed "Local Object Enhanced-UNet". In this figure, the input images are first pre-processed to remove any noise and ensure their suitability for the network's input structures. The UNet performs encoding and decoding operations, generating content-based image segmentation outputs. Each output pixel is assigned an index number that represents the semantic content of the pixel, which enables the identification and extraction of large areas of buildings. By extracting the area, we can specifically focus on the roof regions. The final outputs of this process consist of connected roof regions from each building, which are then ready for automatic roof material analysis.

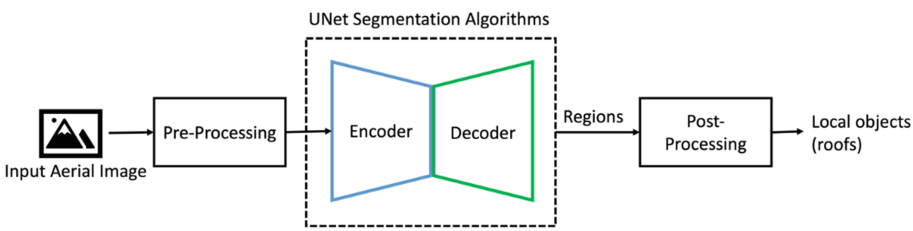


Fig. 1. Local object enhanced UNet system pipeline

Two different neural network frameworks were used to demonstrate the versatility of the proposed algorithm. These frameworks include the Base CNN, which serves as a solid foundation for the algorithm, and the Pre-trained ResNet 50, a well-established and widely used CNN model, which further enhances the performance and robustness of the algorithm. By employing these two distinct frameworks, we showcase the practicality and effectiveness of our proposed algorithm. In the following sections, each component from the system pipeline is introduced.

3.1 The Machine Learning Dataset and Data Pre-processing

The prototype is based on "Semantic segmentation of aerial imagery" [16], which has been published as an open access dataset annotated for a joint project between Humans in the Loop and the Mohammed Bin Rashid Space Centre in Dubai, the UAE. The dataset covers eight different areas and contains five main content categories, namely "building", "land", "road", "vegetation", and "water". Additionally, any other content that does not fall into these categories are labelled as "unlabelled". Figure 2 showcases a selection of samples from the dataset, displaying both the original aerial image and the manually annotated regions used for machine learning purposes.

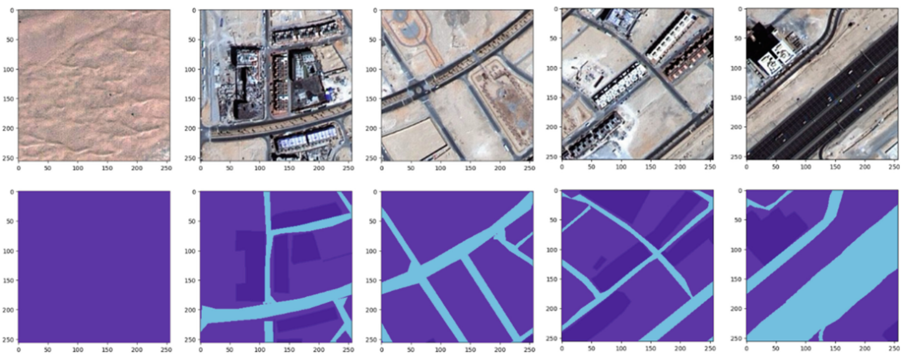


Fig. 2. Samples (top: original image; bottom: ground truth manual annotations) from "Semantic segmentation of aerial imagery"

The original aerial images and their annotation images are all cut into smaller tiles (255 pixels by 255 pixels). Generating smaller tiles is beneficial for several reasons. Firstly, it allows for an increased number of training samples, which allows the accuracy and robustness of machine learning models. Moreover, utilising a smaller tile size helps to reduce the overall data load required for machine learning, particularly during batch training, thereby streamlining and expediting the process. Furthermore, resizing the image into smaller tiles enables it to conform to the input data format of CNN and ResNet, facilitating efficient data training.

3.2 CNN-Based UNet

UNet is an improved CNN deep learning architecture introduced by Ronneberger et al. in 2015 [14], which is regularly utilised for image segmentation tasks. The UNet architecture contains an encoder-decoder structure. The role of an encoder is to capture context and extract attributes from the input image. The UNet encoders comprise convolutional layers with down-sampling procedures to lower the spatial dimensions of the feature maps while increasing the number of channels. The decoder part seeks to reconstruct the segmented image from the extracted features. It consists of up-sampling operations, such as transposed convolutions or interpolation, to progressively expand the spatial dimensions of the feature maps.

3.3 Training Approach

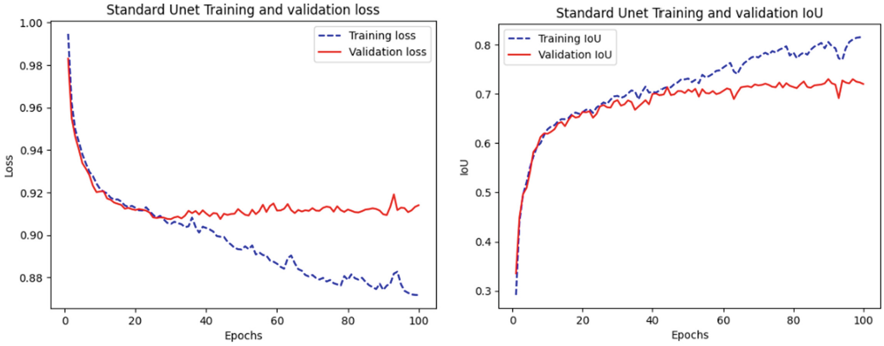
In a CNN structure, all the weighted filters in each feature map are ready to be adjusted through the training process. This adjustment allows the filters to learn and adapt to the specific features and patterns in the input data. To train the UNet model, a commonly used approach is to use a loss function to determine the variation between the predicted segmentation output and the ground truth (i.e., manual annotations of the segmentation), such as binary cross-entropy [17]. Optimization algorithms, such as backpropagation, play a crucial role in this training process. Backpropagation allows the model to iteratively update and adjust the weights in each filter based on the observed errors or discrepancies between the predicted segmentation output and the ground truth.

For instance, Fig. 3(a) demonstrates the learning process of our proposed CNN-based UNet model training from scratch until converged. The loss function gives less loss through the training process and the accuracy of the model showed improvement. In this study, we focused on measuring the segmentation accuracy of the system. This was done using the Intersection over Union (IoU) rate, a metric that is generally applied in image segmentation tasks. The IoU rate measures the overlap between the predicted segmentation mask and the ground truth mask. By calculating the IoU rate for each segmented object or region, we were able to assess the accuracy of our system by identifying and delineating different objects in the aerial images. This evaluation metric helped us understand the effectiveness of our segmentation algorithm and guided us in optimizing it for better accuracy and performance. Figure 3(b) demonstrates the progress of the average IoU rate increases throughout the training process over 100 training iterations (epochs).

3.4 Transfer Learning and ResNet 50-Based UNet

The key behind transfer learning is to transfer the learned features or representations from the pre-trained neural network model to a new model. In practice, transfer learning involves taking a pre-trained model, normally trained on a large dataset such as ImageNet [18], while removing the last few layers that are task specific. The pre-trained layers operate as a feature extractor, capturing common patterns and features that are useful for a wide range of tasks.

In this study, we adapted and adopted the original CNN structure through transfer learning. We selected the ResNet 50 model [19] which has been extensively used for diverse tasks. ResNet 50, short for Residual Network 50, is a deep convolutional neural network architecture, including image classification and object detection. We used a ResNet 50 model that has been pre-trained on the ImageNet dataset.



(a) (b)

Fig. 3. (a) CNN-Based UNet model training and validation loss over 100 training iterations; (b) CNN-Based UNet model training and validation performance gain (IoU score) over 100 training iterations

3.5 Local Object Feature Extraction

One of the disadvantages of using a CNN structure for image segmentation is that it can sometimes overlook small details. This can happen due to the nature of the convolutional neural network architecture, which is primarily designed to capture and process larger patterns and features in an image. It is important to note that in aerial and satellite image segmentation tasks, the lack of detailed information can lead to under-segmented results. While the main region may be recognizable, it becomes challenging to distinguish individual objects within the region, such as each house in a residential area.

To effectively address the problem, we strategically implemented the localised, low-level image edge features. By doing so, we were able to significantly augment and amplify the previously overlooked intricacies and nuances, ultimately resulting in a more comprehensive and refined solution.

As illustrated in Fig. 4, the edge features in this study are comparable to the Canny edge detection algorithms [20], which are based on the gradient of the image. In addition, we used the Mean Shift clustering algorithms to cluster similar colours and pixel locations. This approach may result in the over-segmentation of regions, but it effectively preserved all the essential details, including the precise locations of the regions of interest for roof materials analysis. By incorporating these advanced techniques, we were able to obtain comprehensive and detailed insights into the analysis of roof materials. Since the Mean Shift clusters are unsupervised, one notable advantage is that the local shape has been extracted without requiring manual annotation. As shown in Fig. 4, the detailed building boundaries were not included in the annotation process due to the significant manual workload involved. However, the final output can automatically detect those details and improve the segmentation results.

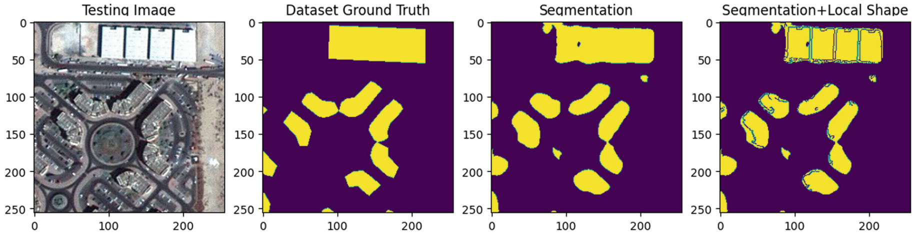


Fig. 4. Sample from output segmentation and local object shape feature extraction

4 Results and Analysis

The section presents a comprehensive evaluation to demonstrate a thorough examination of the segmentation performance using the reliability assessment metric, *i.e.*, Intersection over Union (IoU) for all the classes involved, namely "building", "land", "road", "vegetation", "water", and "unlabelled".

The IoU, recognized as the Jaccard index, is a metric utilised to quantify the accuracy of an image segmentation model. It assesses the overlap between the predicted segmentation and the ground truth. This is done by dividing the area of overlap by the area of union of both sets. The IoU score ranges from 0 to 1, where a higher score shows a better match between the predicted and actual segmentations. In this context, we use the IoU metric because it provides a reliable measure of how well our models are performing in terms of accurately identifying and classifying the different classes within our dataset.

The findings revealed that the average IoU score for the base CNN model stands at 58.4%, while the ResNet pre-trained model exhibited a higher average IoU score of 59.7%. These results are particularly noteworthy considering the complexity of the task, involving the accurate identification and classification of six distinct classes within a dataset comprising over 700 diverse data points.

The performance has been further compared as shown in Fig. 5. In this analysis, we separated the training and validation dataset results from the two proposed methods. After 100 training epochs, it is observed that all the modules have converged. It is worth mentioning that due to the scope of the models, the Base CNN UNet algorithms can reach their top performance earlier compared to ResNet, primarily because of their simpler structure. However, when considering the overall performance on both the training and validation dataset, the pre-trained ResNet outperforms the Base CNN UNet algorithms.

On the other hand, the learning process of the ResNet has some significant jumps during the training. These jumps are primarily caused by using a large learning rate, which can be adjusted and optimised through the iterative machine learning engineering process. By carefully analysing and monitoring the training progress, it is possible to identify these jumps and implement corrective measures to ensure the smooth and consistent learning of the ResNet model.

In addition to the experiments, we conducted further investigations that specifically examined the overall performance of the system by integrating both segmentation and local shape features. It is important to note that the annotation from the public dataset lacked any information regarding the local shape, therefore, to evaluate and discuss the

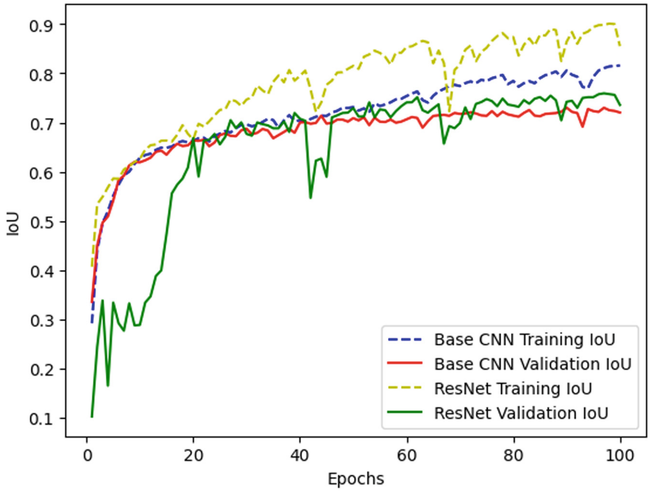


Fig. 5. IoU rate of two different neural networks across training and validation datasets

results obtained, we relied solely on subjective visual assessment using the following examples.

From Fig. 6, it is observed that each sample in the dataset is structured with two rows and four columns. The first rows represent the segmentation algorithms implemented using the Base UNet model, while the second rows correspond to the ResNet-based UNet. The first two columns of each sample contain the original tiles extracted from the dataset and the manual segmentation ground truth, respectively. Moving on to the subsequent columns, they showcase the segmentations obtained through various techniques, with the final columns specifically highlighting the enhanced version of the local shape features.

ResNet is known for its outstanding accuracy and precision in capturing the intricate details of various buildings. It excels in accurately preserving the sizes, proportions, and overall structure of the surrounding regions, demonstrating its exceptional capabilities in the field of image recognition and analysis. Additionally, ResNet has been proven to be highly proficient in maintaining the shapes and intricate details of objects compared to the base CNN version. This attribute makes ResNet a popular choice for tasks that demand a high level of preservation of shapes and fine details.

Focusing on the local shape feature enhancement, by using techniques such as Mean shift and gradient-based shape features, we were able to achieve the fine details from buildings in the local regions. The integration of these methods into our system significantly enhanced its performance in accurately capturing and preserving intricate architectural elements. This allowed us to focus specifically on the buildings and extract their detailed features with precision.

It is also worth mentioning that both algorithms demonstrate a certain level of robustness in dealing with shadows. This is an important characteristic as shadows can often pose challenges in satellite image processing tasks. Shadows can obscure important details and affect the accuracy of segmentation and shape analysis. Therefore, the ability of the algorithms to handle shadows effectively enhances their overall performance

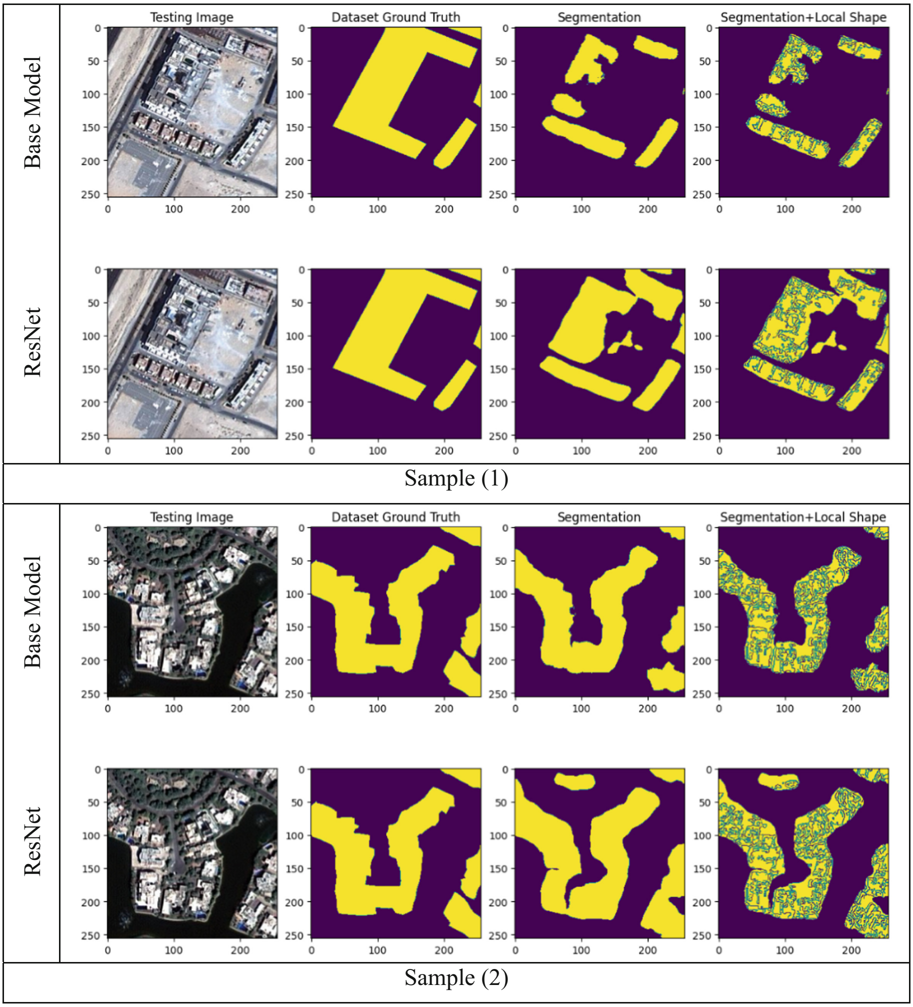


Fig. 6. Outputs (1–2) and side-by-side comparison between two algorithms (Base CNN model and ResNet model)

and reliability. By being robust in shadow handling, the algorithms can provide more accurate and reliable results, leading to improved image analysis and interpretation. This robustness ensures that the algorithms can effectively handle various lighting conditions and produce consistent and accurate outputs, regardless of the presence of shadows.

By integrating these methods into our system, we were able to specifically target the buildings in the images and extract their detailed features. This is crucial because, in urban areas, buildings are often the most prominent and important objects of interest. By enhancing the local shape features, we can effectively isolate the buildings from the surrounding environment and focus solely on extracting their architectural details.

This allows us to acquire a more thorough understanding of the urban landscape and facilitates further analysis and interpretation of aerial photography and satellite images.

The applied methods demonstrate the potential to produce accurate data from aerial imagery. The outputs provide insight into acquiring information on physical parameters and a deeper understanding of generating geo-located building geometry. The study identifies several implications of the applied methods to support initiatives for climate resilience. First, obtaining finer-scale details for improving urban climate simulation as many existing urban climate models lack a high level of detail. Second, providing a reliable tool to enhance climate-adaptation observation in terms of revealing data on how cities align with the climate goals. Third, testing the reliability of aerial imagery and satellite-based observations, the changing behaviour of the same environmental components and land-use activities. Fourth, identifying sources of heat stress, such as the density of built-up areas, percentages of greenery spaces, and aspects of biodiversity. Fifth, opening the possibility of enriching urban analytics with correct data to guide the development of environmental strategies for smart and sustainable cities. Finally, supporting green investments and establishing databases for future scenarios to support city regulators and urban planners.

5 Conclusion

The study demonstrated the efficiency of segmentation algorithms, particularly the Base UNet model and the ResNet-based UNet, in analysing aerial images of urban areas. UNet-based image segmentation was successful in extracting detailed features from buildings with high precision and has proven its effectiveness. The integration of UNet-based segmentation algorithms, including the CNN UNet model and the ResNet-based UNet, has significantly enhanced the image analysis process. In addition, by incorporating techniques such as Mean shift and gradient-based shape features, the study has successfully enhanced the local shape features, enabling the precise capture and preservation of intricate architectural elements. This fine-grained analysis provides a comprehensive understanding of the urban landscape and facilitates further interpretation of the aerial images.

These methods can be seen as extended tools in the field of urban analytics, particularly in urban computing, a promising field that needs further research and development. These methods have specific implementation scenarios in the practice: (1) enabling active technologies such as City Information Modelling (CIM) to develop models for urban digital twins. These methods are capable of presenting correct digital replicas of physical systems or environments to guide in performing real-time monitoring, simulation, and analytics of ecosystems. (2) offering the possibility to provide computational resources to train models in machine learning and AI. (3) guiding the use of computer vision coupled with deep learning to value big data at urban levels. (4) eliminating predictions and advancing towards decision-making to support policymakers to optimise policy.

References

1. Hincks, S., Carter, J., Connelly, A.: A new typology of climate change risk for European cities and regions: principles and applications. Glob. Environ. Chang. **83**, 102767 (2023)
2. Singleton, A.D., Spielman, S., Folch, D.: Urban Analytics. Sage (2017)
3. Omrany, H., Al-Obaidi, K.M., Hossain, M., Alduais, N.A., Al-Duais, H.S., Ghaffarianhoseini, A.: IoT-enabled smart cities: a hybrid systematic analysis of key research areas, challenges, and recommendations for future direction. Discov. Cities **12**(1), 2 (2024)
4. Wang, G., Meng, X.: Evaluation and calculation of urban carbon emission reduction potential based on spatiotemporal geographic model. Int. J. Thermofluids **20**, 100478 (2023)
5. The Alan Turing Institute. https://www.turing.ac.uk/research/research-programmes/urban-analytics, last accessed 2023/12/29
6. Shi, J., Pan, Z., Jiang, L., Zhai, X.: An ontology-based methodology to establish city information model of digital twin city by merging BIM, GIS and IoT. Adv. Eng. Inform. **57**, 102114 (2023)
7. Mashala, M.J., Dube, T., Mudereri, B.T., Ayisi, K.K., Ramudzuli, M.R.: A systematic review on advancements in remote sensing for assessing and monitoring land use and land cover changes impacts on surface water resources in semi-arid tropical environments. Remote. Sens. **15**(16), 3926 (2023)
8. Srivastava, A., Maity, R.: Assessing the potential of AI–ML in urban climate change adaptation and sustainable development. Sustainability **15**(23), 16461 (2023)
9. Guo, Y., Liu, Y., Georgiou, T., Lew, M.S.: A review of semantic segmentation using deep neural networks. Int. J. Multimed. Inf. Retr. **7**, 87–93 (2018)
10. Ghaseminik, F., Aghamohammadi, H., Azadbakht, M.: Land cover mapping of urban environments using multispectral LiDAR data under data imbalance. Remote. Sens. Appl.: Soc. Environ. **21**, 100449 (2021)
11. Sultana, F., Sufian, A., Dutta, P.: Evolution of image segmentation using deep convolutional neural network: a survey. Knowl.-Based Syst. **201**, 106062 (2020)
12. Khryashchev, V., Ivanovsky, L., Pavlov, V., Ostrovskaya, A., Rubtsov, A.: Comparison of different convolutional neural network architectures for satellite image segmentation. In: 2018 23rd Conference of Open Innovations Association (FRUCT), pp. 172–179. IEEE (2018, November)
13. Marri, V.D., P, V.N.R.: RNN-based multispectral satellite image processing for remote sensing applications. Int. J. Pervasive Comput. Commun. **17**(5), 583–595 (2021)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pp. 234–241. Springer International Publishing (2015)
15. Singh, N.J., Nongmeikapam, K.: Semantic segmentation of satellite images using deep-UNet. Arab. J. Sci. Eng. **48**(2), 1193–1205 (2023)
16. Roia Foundation, Semantic segmentation of aerial imagery. https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery, last accessed 2023/12/29

17. Good, I.J.: Rational decisions. J. Roy. Stat. Soc.: Ser. B (Methodol.) **14**(1), 107–114 (1952)
18. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009, June)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
20. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 679–698 (1986)