Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 5, pp. 4023–4031, DOI 10.12694/scpe.v25i5.3155

RESEARCH ON DEEP LEARNING-BASED ALGORITHM FOR DIGITAL IMAGE COMBINATION AND TARGET DETECTION

SHANLU HUANG*AND JIALIN LAI†

Abstract. This study uses deep learning techniques to improve target recognition and digital picture processing, combining efficiency and accuracy in the fields of computer vision and image processing. Different situations, heterogeneous circumstances in the environment, and a wide range of image properties present obstacles for the traditional approaches of combining images and target recognition. To address these issues, our research suggests a novel method that makes use of deep learning methods to identify relevant characteristics and trends from a variety of sources that provide diverse pictures. As part of the research process, a complex deep learning system that can recognize ordered representations of input photos is developed and trained. We will investigate whether faster RCNN are suitable for capturing temporal and spatial relationships in the image data. To maximize the model's performance, deep learning techniques will be used to make use of pre-trained networks on sizable datasets. Benchmark datasets will be used in the method's assessment, and it will be pitted with conventional image processing techniques. The accuracy and dependability of the algorithm's performance will be evaluated using quantitative metrics including precision, recall, and F1-score. Furthermore, qualitative evaluations will be conducted to determine the visual appeal and interpretive capacity of the created composite images.

Key words: deep learning, faster RCNN, digital image combination, target detection, satellite images

1. Introduction. The branch of computer science known as artificial intelligence (AI) studies how to make machines intelligent. In a perfect world, these devices would react similarly to humans in terms of perception, comprehension, and problem-solving decision-making [21, 5, 26]. Artificial Intelligence (AI) encompasses a broad range of fields, most of which are related to the senses that humans have, including computer vision (CV), the processing of natural languages (NLP), oversight, and robots. Through its ability to comprehend digital images and movies, computer vision is a branch of computer science that strives to emulate human vision [16, 2, 27, 15]. It analyses photos using a variety of optimization approaches and techniques. CV is a multidisciplinary field that includes automation, math, probability, artificial intelligence, and recognition of patterns. The branch of artificial intelligence called machine learning (ML) uses data to learn instead of explicit programming [8].

A more intricate and sophisticated model is needed to comprehend pictures and videos. Neural networks (NNs) are remarkably adept at processing vast volumes of data (such as photos and videos) and deciphering it, according to research findings. Scientists were able to resolve challenging issues such picture categorization, recognizing objects, recognition of objects, and segmentation of instances recognition of optical characters by utilizing neural networks in CV. Using deep learning methods, computer vision additionally plays a role in the analysis and detection of objects in images. By resolving the issues, CV has influenced several industries, including the analysis of documents, self-driving cars, medical imagery analysis, and satellite picture research. [14].

For several years, one of the main goals of computer vision research has been to identify objects. The primary objective of recognizing objects is to identify an instance in pictures and videos [28]. Using a bounding box, object identification in CV refers to identifying things of interest (such as people, pets, dogs, cycles, etc.) at a given spot in an image [1]. In the fields of artificial information, machine vision, and robotic seeing, object detection finds numerous uses, such as in augmented reality, security, and surveillance. Two types exist for

*School of Public Administration, Guangxi Technological College of Machinery and Electricity, Nanning 530007, China
†School of Public Administration, Guangxi Technological College of Machinery and Electricity, Nanning 530007, China (jialinlaidigita@outlook.com)

object detection. Finding general categories (person, cat, etc.) is the first form of detection; the second type targets particular examples, like the president's face.

Identifying objects in satellite images is a crucial, essential, and difficult task since objects are small, multi-oriented, and densely grouped. Thus, the main challenge is to identify and locate small objects in satellite images. Because the low-resolution image dataset shortens the training period, we have created a custom dataset with low-resolution images of objects (like small-sized aircraft) to achieve good accuracy with a minimal amount of computational power. Using a custom dataset, we have analysed the speed and accuracy of various object detection pipelines.

The main contribution of the proposed method is given below:

1. We assembled a dataset of satellite photos of airplanes and pre-processed it for training and testing purposes.
2. To speed up the target identification process, the suggested algorithm makes use of the Faster R-CNN architecture, which is well-known for its effectiveness in object detection tasks.
3. The model overcomes the computational performance constraints of standard methods by effectively localizing and classifying targets inside the images using region-based convolutional neural networks.
4. Using a bespoke dataset, the effectiveness of the main algorithms for the identification and categorization of aircraft in satellite imagery was compared in terms of execution speed and accuracy.

The remaining sections of this paper are structured as follows: Section 2 discusses the related research works, Section 3 describes the digital image combination and target detection, Section 4 presents the methods used to adopt the proposed model, Section 5 discusses the experimented results and Section 6 concludes the proposed system with future work.

2. Related Works. In the past few years, deep learning and machine learning approaches have been used to overcome many issues related to object identification in satellite imagery. There are three pipelines that offer real-time remedies: YOLO, SSD, and Faster-RCNN. The cutting-edge real-time object recognition framework YOLO (you only look once) uses a 416 x 416 resolution image and is based on a CNN (convolutional neural network) algorithm [23, 19]. The state-of-the-art framework Faster RCNN uses a 1000×600 resolution image and is based on the region suggestion method. The SSD (single shot detector) architecture operates on either 300×300 or 512×512 pixels per image, extracting feature maps across various layers and applying CNN filters to recognize an item.

The International Society for Photogrammetry and Remote Sensing (ISPRS) Vaihingen and dam benchmark datasets [6], which include high-resolution photos followed by CNN for fine-tuning and hitting state-of-the-art accuracy, were subjected to dense labelling by the author [9, 4, 29]. By using region-based approaches and classification algorithms, the researcher in [20] focused on remote sensing and localization and produced improved object localization outcomes. Nevertheless, the region-based method's significant latency prevented the huge area from being covered (40 s covered area of 1280×1280 pixels). Although it was shown to be slower for segmentation, the author's work [12] used separation and additional processing approaches and produced trustworthy findings regarding automated road detection in satellite data.

Sparse and collaborative representation, as well as kernel-based machine learning, have seen the effective application of machine learning in HTD. By expanding traditional statistical techniques, several kernel-based target detectors have been presented, such as kernel target-constrained interference-minimized filter (KTCIMF) [25], kernel adaptive subspace detector (KASD) [10], and kernel orthogonal sub-space projection (KOSP) [11]. However, a lot of presumptions are also extensively relied upon by these procedures. Regarding limited and cooperative depictions, since the author developed a sparsity-based target detector (STD) [17], several other useful works have been presented. These include the hybrid sparsity and statistics detector (HSSD) [24], the combination of sparse and collaborative representation (CSCR) [13], and the sparse representation-based binary hypothesis-based detector (SRBBHD) [19].

Only a few techniques have been developed for the relatively new use of deep learning to HTD. Guided detectors use synthesizing to primarily increase target data. They then build an end-to-end detector through extensive pixel-pair training. Among the well-liked techniques are two-stream convolutional network-based target detector (TSCNTD) [17], deep network-based HTD (referred to as HTD-Net) [18], and convolutional neural network target detector (CNNTD) [7]. Furthermore, to transfer information from a large-sample domain

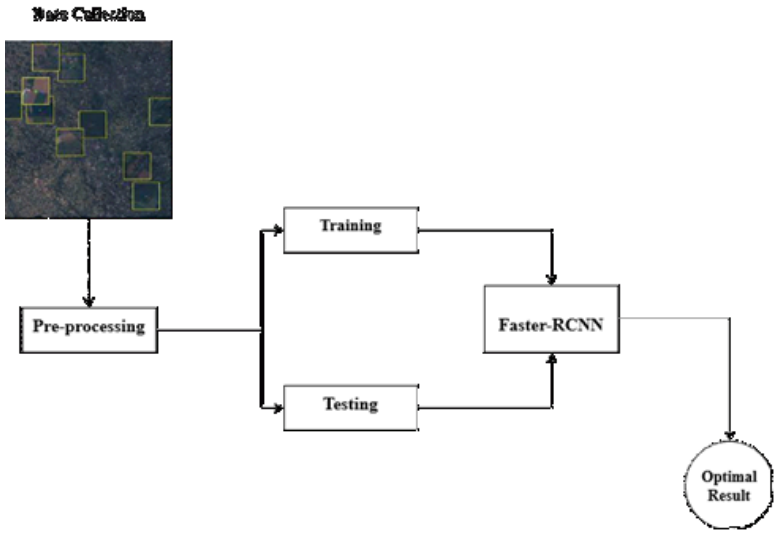

Fig. 3.1: Architecture Diagram of Proposed Method

of origin to a small-sample target domain, a domain adaptable learning model has been developed [22]. Under specific restrictions, unsupervised methods often improve the discriminating ability with unsupervised networks and then use a straightforward matching strategy to detect targets [3].

3. Proposed Methodology. In this work, Faster RCNN method is used for digital image combination and target detection. Initially the dataset is collected, in this work the aircraft satellite images are used as a dataset. Next the collected dataset is pre-processed and then the data is divided for training and testing. The training and testing is carried out by deep learning method Faster-RCNN. In figure 3.1 shows the architecture diagram of proposed method.

3.1. Dataset Collection. One of the most important phases in the entire object identification process is creating the dataset since the dataset has a significant impact on the model's accuracy and performance. It is the most crucial factor to consider while evaluating and examining the efficacy of different algorithms. The internet makes it possible to use larger images in a multitude of categories to accurately depict the intricacy and diversity of objects. The emergence of extensive datasets such as millions of photos has been crucial in facilitating exceptional object detection capabilities.

We obtained satellite images with a 1920 x 1080-pixel resolution using Google Earth. Since they are typically classified, real-time satellite surveillance photographs are extremely difficult to find. For this reason, Google Earth is your best bet when looking for satellite photos of aircraft. Consequently, we tried to locate as many pictures of aircraft as we could. The dataset ought to be larger, but our options are limited. To cut down on training time, we separated the collected photos into 550 $\times$ 350 resolution after collecting. Next, we manually eliminated every picture that didn't include any items. In our dataset, there are 442 photos including 2213 aircraft objects. Next, we tagged photos using the labelling tool.

Pre-trained models are adept at extracting complex features from images, thanks to their exposure to diverse datasets. These features can range from basic textures and shapes to more intricate patterns, providing a rich set of characteristics for the system to use in target recognition tasks. Leveraging pre-trained models can drastically reduce the time and resources needed to train deep learning systems from scratch. Since these models have already learned a broad set of features, they require less data and fewer iterations to adapt to the specific nuances of a new task.

3.2. Pre-processing. An essential first step in getting the data ready for further examination or use is pre-processing satellite photos. To improve the quality, fix distortions, and retrieve pertinent information,

a sequence of actions must be followed. Adjust photometric aberrations unique to each sensor to guarantee uniformity in color and brightness throughout the image. Adjust for distortions in geometry brought on by differences in geography, Earth's curvature, and viewing angles of satellite sensors. To create an accurate planimetric description of the image, this stage entails orthorectification.

Adjust for environmental variables including skies, haze, and particles to enhance the image's quality. For applications involving remote sensing, this is especially crucial. If required, adjust the image's spatial resolution to conform to the analysis's specifications or to match other datasets. This frequently entails resampling methods such as cubic or bilinear convolution.

3.3. Training and Testing the data using Faster-RCNN. The Faster RCNN is a two-stage detecting architecture that involves the categorization and localisation of objects in the second phase and the creation of areas in the first. Fast RCNN has a quick detection process and is dependent on external region recommendations. According to recent research, CNN can localize items in CONV (convolutional) the layers, but its performance is less in fully linked layers. Consequently, a targeted quest for generating regional proposals took the place of CNN. They suggested replacing selective search with a precise and effective region proposal network (RPN) to generate region proposals. They split the structure into two components: fast RCNN for object categorization and the localization of operations and RPN for region proposal creation.

Design or utilize existing APIs (Application Programming Interfaces) that allow for smooth data exchange and communication between the deep learning system and existing software. This may involve developing custom middleware or adapters. Ensure that data formats, including input images and output recognition results, are standardized across systems to facilitate easy sharing and processing.

The categorization and placement of objects using bounding boxes is carried out by an extensive number of convolutional layers used in RPN and the last convolutional layers in the faster RCNN. Figure 4.2 shows the network topology of the faster RCNN. When features are retrieved by CONV layers, RPN creates k n $\times$ n anchor boxes with varying aspect ratios and sizes. Every n $\times$ n anchor is transformed into a low dimensions vector, like 512 for the group known as Visual Geometry Group (VGG) and 256 for ZF. These vectors are then fed into two fully linked layers, which comprise layers for object categorization and bounding box regressors.

Since RPN is a sort of fully convolutional network, it shares features with the rapid RCNN and facilitates the computing of region suggestions efficiently. Instead of using manually created features, CNN uses faster RCNN only for feature extraction (Figure 3.2). Using three hundred suggestions per image, the faster RCNN with the VGG16 model reaches object detection accuracy on the PASCAL VOC dataset at 5 frames per second on the GPU. The author investigated the role of region suggestion the generation through selective search and region proposal generation through CNN considering the quicker RCNN growth. They concluded that CNN-based RPN includes less geometric data for identifying objects in the CONV Layers as opposed to FC layers.

$$(himg, wimg, x, y, w, h, objectives)$$

K is generated at each sliding window location when training a faster RCNN with anchors and various proposals. A class probability of object or not object is represented by a 2K score in the CLS layer, whereas the 4K boxes with coordinates in the Reg layer. K anchors, sometimes known as boxes, are the subject of the K parameter. They produce nearly WHK anchors at the convolutional feature map W $\times$ H by using k = 9 with three scales and three aspect ratios at each sliding window. To solve the multiscale problem based on anchors, Faster RCNN employs CNN for features computed on a single scale image. Sharing features and addressing multiscale at a lower cost are two advantages of this.

They give each object a binary label that indicates whether the object is present or absent for training purposes. They give anchors a favourable label. Anchors can be computed in two different methods. Ground truth boxes assign labels to numerous anchors. Initially, one picks those anchors whose crossover over union is high with a ground truth box. Secondly, one selects those anchors whose intersection over union is bigger than 0.7 with a ground truth box. As a result, the second requirement is inappropriate for accurate anchor prediction. As a result, they apply the initial criterion, which gives anchors positive labels and has the highest IOU with the ground truth box.

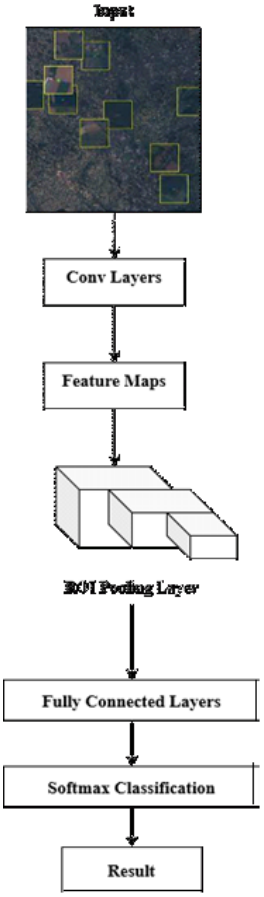

Fig. 3.2: Overall Process of Proposed Method

4. Result Analysis. Our work concentrated on creating and refining a Faster R-CNN-based algorithm for target detection in satellite photos of aircraft and digital image combining. When compared to traditional methods, the suggested algorithm showed notable improvements in terms of speed, accuracy, precision, recall and F1-score.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} X100 \tag{4.1}$$

$$precision = \frac{TP}{TP + FP} X100 \tag{4.2}$$

$$recall = \frac{TP}{TP + FN} \tag{4.3}$$

When evaluating the accuracy of Faster R-CNN target recognition on satellite images, the model's predictions are usually compared with ground truth annotation. Determine how many times the model's predictions coincide with the actual data (accurately predicted targets). Find out how many targets there are in the dataset overall. Utilize the formula to determine the accuracy. Remember that although while accuracy is a widely used metric, it might not be enough in all situations, particularly when working with datasets that are unbalanced

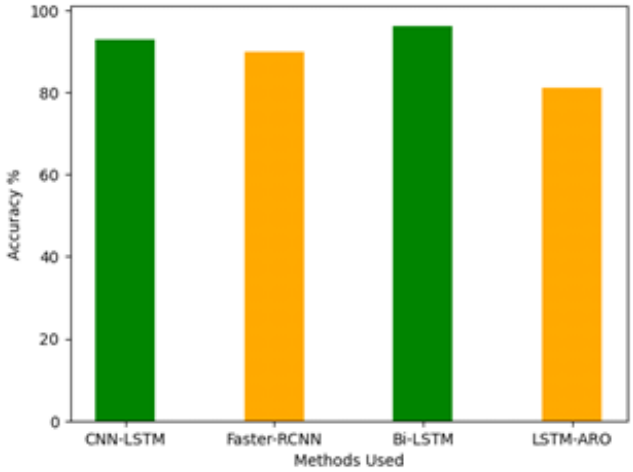

Fig. 4.1: Accuracy

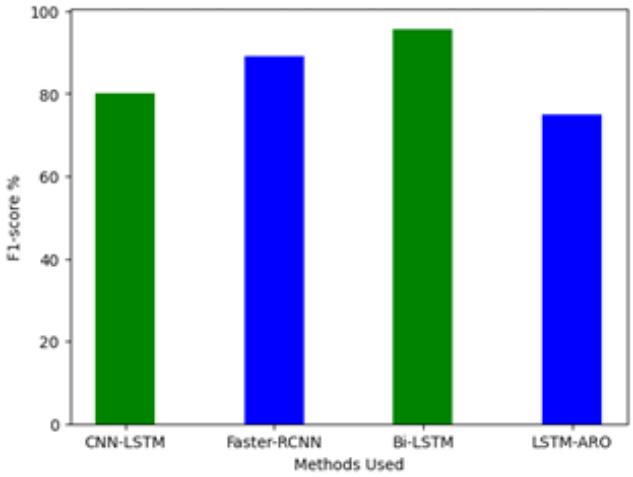

Fig. 4.2: F1-Score

or in situations where false positives or false negatives have distinct outcomes. In these circumstances, further measures such as recall, precision, and F1-score may be considered to offer a more thorough assessment of the model's effectiveness in target detection on satellite photos. In figure 4.1shows the evaluation of Accuracy.

A high F1-score suggests a good trade-off between precision and recall when evaluating a Faster R-CNN model for satellite image target recognition, indicating that the algorithm is successfully locating and recognizing targets in the images. When analysing F1-score results, it's crucial to consider the requirements of the application as well as the implications of false positives and false negatives. In figure 4.2 shows the evaluation of Precision. When employing Faster R-CNN for object detection tasks in satellite pictures, precision is an essential evaluation criterion. By quantifying the precision of the model's positive predictions, one can ascertain the proportion of projected positive instances that turn out to be true positives. figure 4.3 shows the evaluation.

Recall is a crucial parameter in satellite image analysis that assesses the model's accuracy in identifying and capturing all pertinent instances of the target class in the dataset when employing a Faster R-CNN-based algorithm. Recall is critical in the context of satellite photography since it offers insights into the algorithm's

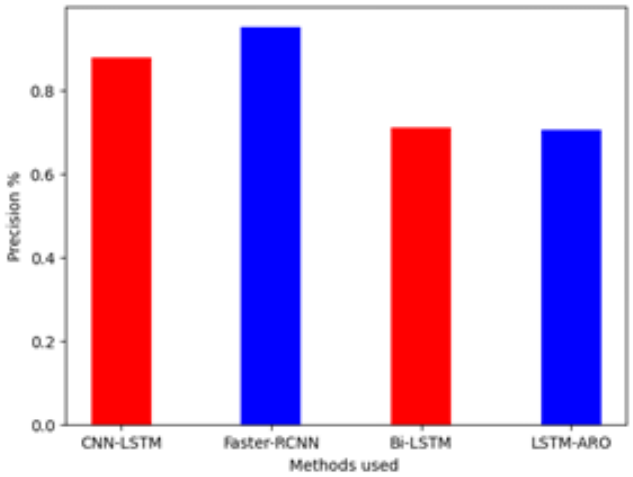

Fig. 4.3: Precision

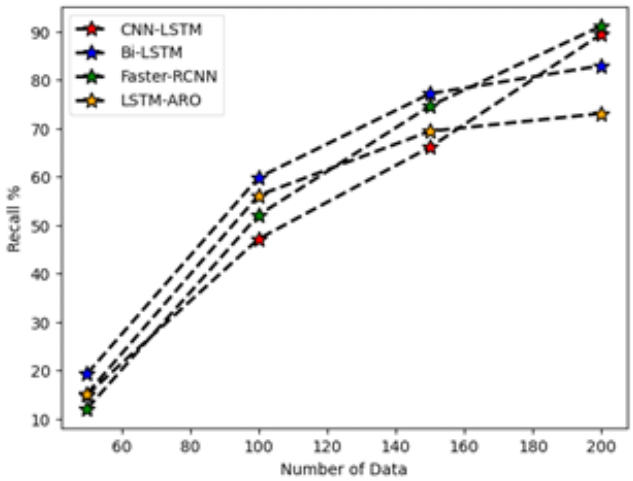

Fig. 4.4: Recall

capacity to identify every instance of the target class, guaranteeing that no significant data is overlooked. A high recall score means that the model can identify most real instances of the target class and effectively reduce false negatives. In figure 4.4 shows the evaluation of Recall.

5. Conclusion. To combine efficiency and accuracy in the domains of computer vision and image processing, this work employs deep learning approaches to enhance target detection and digital picture processing. Traditional ways of merging images and target recognition face challenges from varying scenarios, variable environmental conditions, and a broad range of image attributes. The study we conducted proposes a novel approach to address these problems by using deep learning techniques to extract significant features and patterns from several sources that offer a range of images. A sophisticated deep learning system that can identify ordered presentations of input photographs is created and trained as part of the study process. We will examine if temporal and geographical links in the visual data can be captured by quicker RCNN. Deep Learning techniques will be applied to utilize pre-trained networks on large datasets to optimize the model's performance.

The method will be evaluated using benchmark datasets and compared to traditional image processing techniques. Quantitative measurements like precision, recall, and F1-score will be used to assess the algorithm's correctness and reliability. Additionally, qualitative assessments will be carried out to ascertain the composite images' visual appeal and interpretive potential.

Acknowledgement. This work was sponsored in part by the Basic Ability Improvement Project of Young and middle-aged people in Guangxi - Research on the integrated development path of "Intangible cultural heritage + Cultural Innovation" from the perspective of Rural Revitalization - taking Bobai miscanthus weaving as an example (2023KY1102).

REFERENCES

[1] Q. Ali, M. J. Thaheem, F. Ullah, and S. M. Sepasgozar, *The performance gap in energy-efficient office buildings: how the occupants can help?*, Energies, 13 (2020), p. 1480.
[2] G. Callebaut, G. Leenders, J. Van Mulders, G. Ottoy, L. De Strycker, and L. Van der Perre, *The art of designing remote iot devices—technologies and strategies for a long battery life*, Sensors, 21 (2021), p. 913.
[3] S. Chakraborty, J. Phukan, M. Roy, and B. B. Chaudhuri, *Handling the class imbalance in land-cover classification using bagging-based semisupervised neural approach*, IEEE Geoscience and Remote Sensing Letters, 17 (2019), pp. 1493–1497.
[4] S. I. Khan, Z. Qadir, H. S. Munawar, S. R. Nayak, A. K. Budati, K. D. Verma, and D. Prakash, *Uavs path planning architecture for effective medical emergency response in future networks*, Physical Communication, 47 (2021), p. 101337.
[5] Y. Li, Y. Shi, K. Wang, B. Xi, J. Li, and P. Gamba, *Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery*, IEEE Transactions on Image Processing, 31 (2022), pp. 1418–1432.
[6] M. U. Liaquat, H. S. Munawar, A. Rahman, Z. Qadir, A. Z. Kouzani, and M. P. Mahmud, *Sound localization for ad-hoc microphone arrays*, Energies, 14 (2021), p. 3446.
[7] S. Low, F. Ullah, S. Shirowzhan, S. M. Sepasgozar, and C. Lin Lee, *Smart digital marketing capabilities for sustainable property development: A case of malaysia*, Sustainability, 12 (2020), p. 5402.
[8] Manju, P. Bhambu, and S. Kumar, *Target k-coverage problem in wireless sensor networks*, Journal of Discrete Mathematical Sciences and Cryptography, 23 (2020), pp. 651–659.
[9] A. Maqsoom, B. Aslam, M. E. Gul, F. Ullah, A. Z. Kouzani, M. P. Mahmud, and A. Nawaz, *Using multivariate regression and ann models to predict properties of concrete cured under hot weather*, Sustainability, 13 (2021), p. 10164.
[10] H. S. Munawar, *Flood disaster management: Risks, technologies, and future directions*, Machine Vision Inspection Systems: Image Processing, Concepts, Methodologies and Applications, 1 (2020), pp. 115–146.
[11] H. S. Munawar, A. W. Hammad, S. T. Waller, M. J. Thaheem, and A. Shrestha, *An integrated approach for post-disaster flood management via the use of cutting-edge technologies and uavs: A review*, Sustainability, 13 (2021), p. 7925.
[12] H. S. Munawar, H. Inam, F. Ullah, S. Qayyum, A. Z. Kouzani, and M. P. Mahmud, *Towards smart healthcare: Uav-based optimized path planning for delivering covid-19 self-testing kits using cutting edge technologies*, Sustainability, 13 (2021), p. 10426.
[13] H. S. Munawar, S. I. Khan, Z. Qadir, A. Z. Kouzani, and M. P. Mahmud, *Insight into the impact of covid-19 on australian transportation sector: An economic and community-based perspective*, Sustainability, 13 (2021), p. 1276.
[14] H. S. Munawar, M. Mojtahedi, A. W. Hammad, M. J. Ostwald, and S. T. Waller, *An ai/ml-based strategy for disaster response and evacuation of victims in aged care facilities in the hawkesbury-nepean valley: A perspective*, Buildings, 12 (2022), p. 80.
[15] H. S. Munawar, S. Qayyum, F. Ullah, and S. Sepasgozar, *Big data and its applications in smart real estate and the disaster management life cycle: A systematic analysis*, Big Data and Cognitive Computing, 4 (2020), p. 4.
[16] H. S. Munawar, F. Ullah, S. I. Khan, Z. Qadir, and S. Qayyum, *Uav assisted spatiotemporal analysis and management of bushfires: A case study of the 2020 victorian bushfires*, Fire, 4 (2021), p. 40.
[17] H. S. Munawar, F. Ullah, S. Qayyum, S. I. Khan, and M. Mojtahedi, *Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection*, Sustainability, 13 (2021), p. 7547.
[18] Z. Qadir, S. I. Khan, E. Khalaji, H. S. Munawar, F. Al-Turjman, M. P. Mahmud, A. Z. Kouzani, and K. Le, *Predicting the energy output of hybrid pv–wind renewable energy system using feature selection technique for smart grids*, Energy Reports, 7 (2021), pp. 8465–8475.
[19] Z. Qadir, A. Munir, T. Ashfaq, H. S. Munawar, M. A. Khan, and K. Le, *A prototype of an energy-efficient maglev train: A step towards cleaner train transport*, Cleaner Engineering and Technology, 4 (2021), p. 100217.
[20] M. A. Shaukat, H. R. Shaukat, Z. Qadir, H. S. Munawar, A. Z. Kouzani, and M. P. Mahmud, *Cluster analysis and model comparison using smart meter data*, Sensors, 21 (2021), p. 3157.
[21] A. Tahir, H. S. Munawar, J. Akram, M. Adil, S. Ali, A. Z. Kouzani, and M. P. Mahmud, *Automatic target detection from satellite imagery using machine learning*, Sensors, 22 (2022), p. 1147.
[22] J. Theiler, A. Ziemann, S. Matteoli, and M. Diani, *Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation*, IEEE Geoscience and Remote Sensing Magazine, 7 (2019), pp. 8–30.

[23] F. Ullah, *A beginner's guide to developing review-based conceptual frameworks in the built environment*, Architecture, 1 (2021), pp. 5–24.
[24] F. Ullah and F. Al-Turjman, *A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities*, Neural Computing and Applications, 35 (2023), pp. 5033–5054.
[25] F. Ullah, S. Khan, H. Munawar, Z. Qadir, and S. Qayyum, *Uav based spatiotemporal analysis of the 2019–2020 new south wales bushfires. sustainability 2021, 13, 10207*, 2021.
[26] F. Ullah, S. Qayyum, M. J. Thaheem, F. Al-Turjman, and S. M. Sepasgozar, *Risk management in sustainable smart cities governance: A toe framework*, Technological Forecasting and Social Change, 167 (2021), p. 120743.
[27] F. Ullah, S. Sepasgozar, F. Tahmasebinia, S. M. E. Sepasgozar, and S. Davis, *Examining the impact of students' attendance, sketching, visualization, and tutors experience on students' performance: A case of building structures course in construction management*, Construction Economics and Building, 20 (2020), pp. 78–102.
[28] F. Ullah and S. M. Sepasgozar, *Key factors influencing purchase or rent decisions in smart real estate investments: A system dynamics approach using online forum thread data*, Sustainability, 12 (2020), p. 4382.
[29] F. Ullah, S. M. Sepasgozar, M. J. Thaheem, C. C. Wang, and M. Imran, *It's all about perceptions: A dematel approach to exploring user perceptions of real estate online platforms*, Ain Shams Engineering Journal, 12 (2021), pp. 4297–4317.

Edited by: Rajanikanth Aluvalu
Special issue on: Evolutionary Computing for AI-Driven Security and Privacy: Advancing the state-of-the-art applications
Received: Jan 6, 2024
Accepted: Feb 9, 2024