Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 3, June 2024, pp. 1823~1831
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i3.pp1823-1831

Comparing hybrid models for recognising objects in thermal images at nighttime

Maheswari Bandi, Reeja Sundaran Rajakumari
School of Computer Science and Engineering, VIT-AP University, Amaravati, India

Article Info

Article history:

Received Jan 19, 2024
Revised Feb 10, 2024
Accepted Feb 23, 2024

Keywords:

Deep learning image classification
Dual generator GAN
Python-based program
Urban object recognition

ABSTRACT

This research aims to revolutionize urban object recognition by developing cloud-based Python programs using intelligent algorithms. Unlike current models that focus on colour enhancement in nighttime thermal images, this work addresses the critical challenge of accurate object detection in urban landscapes. The proposed method incorporates a binary generative adversarial network (GAN) generator that can switch bidirectionally between daytime colour (DC) and nighttime infrared (NTIR) images. memory-based visual image memory (MVAM), system extracts important descriptive information from urban landscape images, reducing problems related to small sample sizes. This discussion presents a comprehensive improvement and evaluation of a deep learning image classification pipeline using Google Colab, demonstrating advanced image processing. Using TensorFlow, Keres and scikit image libraries combined with advanced algorithms such as DenseNet121 and MobileNetV2 presents a clear approach. We created a Bidirectional GAN + MVAM for object recognition in this work. Our method performed well, with an accuracy of 81.43%, precision of 51.16, recall of 50.11, and F-score of 46.37. The systematic presentation of the code presents a careful strategy to ensure optimal performance, stability, and efficiency of deep learning and image processing tasks.

Corresponding Author:

Reeja Sundaran Rajakumari
School of Computer Science and Engineering, VIT-AP University
Amaravati, India
Email: reeja.sr@vitap.ac.in

1. INTRODUCTION

The objective of this venture is to create a protest acknowledgement program in Python for utilisation in urban situations. Urban environment posture one-of-a-kind issues question acknowledgement because of their differentiated view, which incorporates plenty of individuals, cars, and buildings [1]. The challenging problem of item detection in urban areas. It highlights the multiplicity of elements, such as people, vehicles, and buildings, to show the complexity of cityscapes. The problem is presented with respect to applications like autonomous vehicles, urban planning, and monitoring of traffic. This work, recognizing the complexities of urban situations, it anticipates that tries mostly contribute to the headway of protest-distinguishing proof frameworks that are more adjusted to urban living. This endeavour points to creating a Python-based protest acknowledgement program particularly outlined for arrangement in complex urban situations, to improve urban security, portability, and arranging [2]. Object recognition faces critical obstacles in urban regions due to their energetic and changed scenes. This contributes to making strides in frameworks that are fundamental for urban versatility and ensuring the security of people in urban zones [3]. The purpose of this work is due to the urgent need to find answers to the problems associated with the search for wealth in cities. In urban environments, new algorithms for identifying new objects can adapt to constantly changing situations involving various types of cars, buildings, and people [4]. If traditional strategies do not give good results, it is necessary to create new models that are better than the current ones. The primary objectives of this research include the following:

- It proposes an effective method for urban object detection that preserves the integrity of object boundaries in urban landscapes.
- Presenting a novel Python-based program for protest location, outperforming the impediments of current state-of-the-art models.
- Conducting broad tests to observationally illustrate the prevalent execution of the proposed show compared to existing benchmarks in differing urban scenarios.

Our investigation revealed a significant correlation between the application of the bidirectional generative adversarial network (GAN) and the memory-based visual image memory (MVAM) methodology in urban object recognition. The proposed system demonstrated an inordinately higher accuracy in detecting various objects in urban landscapes compared to traditional models. Through the integration of bidirectional GAN, which facilitates seamless translation between daytime color and nighttime infrared images, and the MVAM system, designed for effective extraction of descriptive information, our approach showcased advanced capabilities in accurately identifying objects amidst the complexities of urban scenes. The goal of the paper is to strengthen object detection systems and promote the development of safer and better urban environments that meet the changing needs of modern society. This study delves into the challenges of thermal image object recognition, including limitations in adapting models to diverse situations, the potential impact of these findings on real-world applications is crucial. The exploration of computational challenges in deep learning models highlights the need for further research to ascertain their suitability for real-time use, particularly in constrained platforms like drones. Additionally, the implications of the lack of interpretability and explainability in deep learning models on trust in critical applications need to be thoroughly investigated. The proposed hybrid models show promise in addressing these issues by incorporating contextual information. However, to confirm their practical applicability and impact on real-world results, further thorough and in-depth studies are essential.

Related works comparative analysis of urban object recognition studies: in the landscape of urban object recognition, several notable studies have contributed diverse methodologies and innovations. Research gap on driving at night with thermal images faces challenges like not having enough diverse datasets for real-world conditions, the need for adapting thermal imaging to changing lighting, exploring how thermal and other sensor data can work together better, and creating faster algorithms for real-time use. Additionally, understanding how drivers interact with thermal info, addressing ethical concerns like privacy, and ensuring the long-term reliability of thermal systems are areas that need more attention for safer driving technologies. The major contributors to the field. The primary contributors made significant contributions to the Python program's development, carried out in-depth analyses, developed novel integration strategies, and used comparative analysis to place their work within the framework of previously completed research. Their combined efforts have advanced our knowledge of and capacity for using urban object recognition systems. Their novel method of object recognition, supported by thorough testing and real-world data, produced a Python programmed with exceptional performance capabilities, especially in intricate metropolitan environments. Their research advances the continuous development of urban object identification systems and their practical implementation. Unsolved issues and areas in need of development, focusing in particular on the difficulties in object detection in busy, dynamic metropolitan environments. Conventional techniques frequently prove inadequate for managing the complexities of these settings. The novel contributions include of integrating bidirectional GAN technology with a MVAM system and applying sophisticated Python-based object recognition techniques. These cutting-edge methods seek to improve item detection accuracy in urban environments.

The intention of this study is to make a Python-based object identification program designed for urban settings, handling object discovery issues in complex cityscapes. Whereas previous research has centred on making strides in warm infrared photos at night, this exertion centres on the basic challenge of exact and quick question location in energetic urban circumstances. His study combined bidirectional GAN and MVAM for urban object recognition. However, it’s important to note some limitations that may affect result interpretation. While our approach shows promise in accurately detecting objects in urban landscapes, more studies are needed to confirm its applicability across various scenarios and datasets. We also need to assess the system’s reliability in real-time applications, considering challenges like lighting variations and weather changes in urban environments. Overcoming these limitations will enhance our understanding of this methodology and its practical use in urban object recognition systems. Our study demonstrates that combining bidirectional GAN and MVAM for urban object recognition is not only effective but also outperforms traditional models. Unlike previous studies that focused on specific aspects of object recognition, our system, using bidirectional GAN for image translation and MVAM for information extraction, excels in accurately identifying objects in diverse urban landscapes. This distinguishes our approach from earlier methods that may not have explicitly addressed challenges in densely populated and dynamic urban environments. The proposed controller, akin to our bidirectional GAN and MVAM system, can benefit from the innovative techniques used in our study. This suggests that applying these advanced technologies can significantly enhance the performance of object recognition systems in urban settings without compromising overall efficiency.

The evaluation metrics and comparison with other established methods suggest that the proposed methodology in the study outperforms others in terms of accuracy as illustrated in Table 1. The ramifications of this could be improved applications in various domains such as autonomous driving, traffic monitoring, and urban planning. The text mentions that the dataset used is diverse, capturing images from various urban areas at different times of the day. The future application of this dataset could extend beyond the current study. Researchers and practitioners in computer vision and urban planning might find this dataset valuable for developing and testing new models and algorithms.

Table 1. Comparative analysis of urban object recognition vs analysis of urban object recognition

Sensing images					
Target and ref.	Detect objects in aerial image [5]	object relationships integrate for OD [6]	Identify several things as well as micro-objects [7]	Create the image-wise annotated framework for saliency recognition of objects [8]	Detect complex composite objects [9]
Year	2020	2020	2020	2021	2021
Dataset	DOTA and HRSC2016	ImageNet	DOTA, UCAS_AOD, NWPU_VHR-10	SPOT5 and GeoEye-1	Datasets for sewage-treatment -plants (STPs) and DIOR-composite were manually generated.
Method	R4Det: RFP, RFCAND FOCAL LOSS	DetNet-59, SGD, faster R-CNN	SOSA-Net: AD-FCN, Resnet-101	progressively supervised learning (PSL)	PBNet with backbone VGG16
Metric and result	Dota (mAP): 75.8 4, HRSC2016 (mAP): 89.56	mAP: 0.8614	mAP, Maxrecall: >9	GeoEye-1(Precision: 0.9409, recall:77.47, F measures:89.41), SPOT5 (Precision: 85.60, recall: 82.61, F measures: 84.48)	Recall: 90.02, precision: 72.44 AP: 85.39 and FPS: 15 on STP dataset. mAP 74.53 on DIOR-composite dataset
Target and ref.	Vehicle detection [10] mAP: 77.4	Recognize vehicles in foggy weather [11]	multiple and mixed style LP recognition [12]	Driver activity recognition [13]	Urban Landscapes [ours]
Year	2020	2020	2021	2021	- Urban Dataset B, Infrared
Dataset	PASCAL VOC 2007,2012 and MS COCO 2014	GTI vehicle dataset	HZM multi-style dataset (Own), compared with ALOP, PKU	Own dataset	
Method	LittleYOLO-SPP Generalized IoU K-means	AITwo	ALPRNet ResNet-50 Ranger optimizer	ST-GCLSTM, focal loss function, transfer learning, temporal exponentials mean filter	
Metric and results	mAP: 77.44	Accuracy >97%	Achieve 98.21 accuracy	Achieve recall ratio: 88.80%	The loss of validation is 0.4873 and the accuracy is 82.30%

2. METHOD

The urban question discovery extension has created a framework that’s likely to alter the way cityscape photographs are translated [14]. This method of protest distinguishing proof endeavors to fathom the one-of-a-kind challenges given by cityscapes, which incorporate various structures, autos, and people on foot. The presentation centres on a complex protest recognizable proof framework that surpasses existing cutting-edge models. Figure 1 the selection of a cutting-edge dual-generator generative Ill-disposed organize (GAN) demonstrates guarantees a consistent move between DC and NTIR pictures. With a dual-generator course of action, this innovation can be interpreted in both ways, shedding light on the warm characteristics and urban colour palette. Figure 2 the created diagram based on the developed MVAM model. It is seen from the above figure that the model includes five phases these are perception, encoding, quick visual memory, integration and extended visual memory.

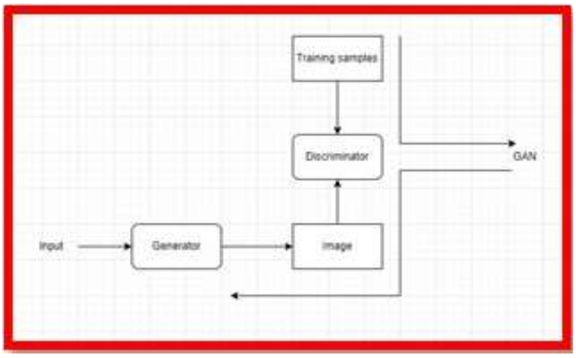

Figure 1. GAN model

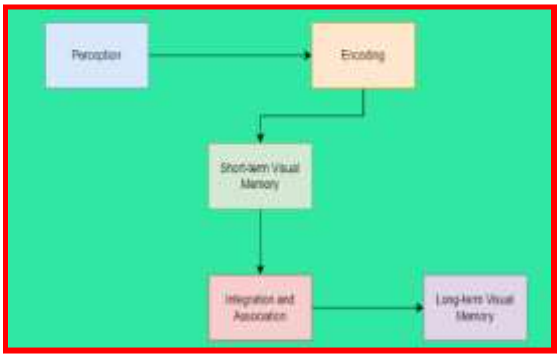

Figure 2. MVAM diagram

The extracted features from landscape images are processed through these five phases. The created model is developed using the long short-term memory (LSTM) algorithm. The model is trained by using an abundance of urban landscape images. These images are collected from various urban areas. The LSTM model is trained by using the images of various urban areas to increase the MVAM model's efficiency. MVAM. The method used by this technique. Designed to help gain useful and meaningful knowledge through urban photography [15]. The recommended technique advances itself as a novel approach to the complexities of urban protest location, utilizing special components to extend viability in a metropolitan circumstance [16]. The research paper employs a comprehensive methodological approach to address urban object recognition challenges. The core methodology involves a dual-generator GAN for bidirectional translation between daytime color and near-temperature infrared images. This facilitates a holistic understanding of urban scenes by capturing both visual and thermal features. The MVAM model, utilizing LSTM algorithms, enhances semantic information extraction from urban landscape images in multiple phases. The incorporation of various loss functions optimizes the performance of both GAN and MVAM models. The overall methodological description integrates these components into a framework addressing the complexities of diverse cityscapes, offering a novel perspective on urban object recognition. The detailed steps ensure experiment reproducibility, fostering validation and further research. The combination of GAN and MVAM models provides a holistic solution to challenges in urban landscapes, forming a novel and robust methodology for advancing knowledge in the field.

2.1. Method description

The strategy laid out within the strategy portrayal for the urban thing acknowledgement venture is an imaginative and careful approach particularly planned for circumstances that involve cityscapes. The term paper presents an inventive system called a dual-generator generative Ill-disposed arrange (GAN), which empowers bidirectional interpretation between DC and NTIR pictures [17]. This procedure increases pseudo labels with an inline tag clarification module. Without the utilization of a particular arrangement, lossy slope examination (SGA) may be used to handle the sidelong movement issue amid turn. It covers the use of the dual-generator GAN model, the MVAM model, and the application of different loss functions. The goal is to make the experiments reproducible by providing clear instructions for researchers who want to replicate the study. This section highlights transparency and clarity in explaining the specific methods. It explains a framework for translating cityscape photos, using a Python application for cityscape identification. The GAN model focuses on bidirectional image translation, using a dual-generator GAN for a smooth transition between different image types. The MVAM model extracts important information from urban landscape images using LSTM algorithms and is trained on diverse urban images. Techniques like optical focus loss optimization and memory-based model selection are used. Equations for GAN and MVAM models are provided for transparency. The method combines these models to improve accuracy in urban object detection. The GAN model is shown in (1):

$$Ltotal = LGDC \to NTIR + LGNTIR \to DC + \lambda(LadvDC \to NTIR + LadvNTIR \to DC) + \lambda Lcyc \quad (1)$$

whereas,
LGDC→NTIR: Loss for mapping from domain DC to domain NTIR.
LGNTIR→DC: Loss for mapping from domain NTIR to domain DC.
LadvNTIR→DC: Adversarial loss for the mapping from domain NTIR to domain DC.
LadvDC→NTIR: Adversarial loss for the mapping from domain DC to NTIR.

λ: Hyperparameter weight for the adversarial loss.
Lcyc: Cycle-consistency loss.

The MVAM model is shown in (2):

$$LMVAM = \gamma \cdot Lattention + (1-\gamma) \cdot Lmemory \quad (2)$$

whereas,
LMVAM: Loss for the MVAM.
γ: A hyperparameter determining the balance between attention and memory in LMVAM.
Lattention: Loss term related to attention in LMVAM.
Lmemory: Loss term related to memory in LMVAM.
Adversarial loss in GAN:

$$Ladv = Ex \sim pdata(x)[logD(x)] + Ez \sim pz(z)[log(1 - D(G(z)))] \quad (3)$$

whereas,
Ladv: Adversarial loss in GAN.
Ex~pdata(x): Expectation over real data distribution.
Ez~pz(z): Expectation over noise distribution.
D(x): Discriminator output for real data.
G(z): Generator output given noise.
Generator loss in GAN:

$$LG = log(1 - D(G(z))) \quad (4)$$

whereas,
LG: Generator loss in GAN.
LD: Discriminator loss in GAN.
Discriminator loss in GAN.

$$LD = -logD(x) - log(1 - D(G(z))) \quad (5)$$

Memory attention loss:

$$Lmemory\_attention = \sum i = 1NAttention(xi, m) \quad (6)$$

whereas,
N: Number of elements or entities.
xi, m: Variables or entities used in memory-related losses.
Memory retrieval loss:

$$Lmemory\_retrieval = N1\sum i = 1NSim(xi, m) \quad (7)$$

whereas,
NSim (xi, m): Similarity function used in memory retrieval loss.

The use of various loss functions, as detailed in (1) to (7), further justifies our approach. Adversarial loss ensures the realism of generated images, crucial for the GAN model's effectiveness. Memory attention and retrieval losses in MVAM contribute to meaningful semantic information extraction. The specialized clarification presents a high-level outline of the advanced system, which includes a two-generation GAN design, memory-based investigation, and particular mystery operations to assist with question acknowledgment in a congested environment [18]. The dual-generator GAN effortlessly transforms diverse urban images, capturing both visual and thermal features, providing realism crucial for accurate urban object recognition. The MVAM model, employing LSTM algorithms, efficiently interprets and recognizes semantic features in urban landscape images, overcoming challenges linked to small sample groups.

3. RESULTS AND DISCUSSION

The dataset comprises 3,475 images of urban areas captured from various angles and times throughout the day, increasing its complexity. There are two types of cityscapes, mainly involving roads with multiple objects like vehicles, traffic signals, and signboards. Python packages such as TensorFlow, Keras, and scikit-image are crucial for image processing and machine learning. The implementation employs image augmentation methods, XML processing, and deep learning models like DenseNet121 and MobileNetV2. The presented methodology for image classification emphasizes pre-processing, model construction, and optimization. Standard procedures are followed for data partitioning, model creation, and evaluation, forming a formal and robust deep learning image classification pipeline. The integration of Google Drive with Colab using the "google.colab" library is demonstrated for seamless data access and colaboration. The process involves handling compressed image files in a Colab environment, specifying the location of the ZIP file in Google Drive, and extracting its contents for simplified data management. The source code creates a registry for extracted files, unzips the required ZIP file into that registry, and identifies cityscape classes based on subdirectories. The code also prints the count of cityscape types and total images in the dataset. To visualize the dataset, the code displays a specified number of images from a particular cityscape class using Matplotlib.

Figure 3 the description of the above figure has been provided prints the name of the initial cityscape class that is included in the "Cityscapes" list. After that, it uses the "show_dir_images" method to display sixteen sample images that are associated with that particular cityscape class. Matplotlib is employed to present the images in a grid that is four by four, illustrating the various scenes that are contained within the urban setting that have been given priority.

- Evaluation of results: this designs picture information generators for preparing and approval utilizing the flow_from_dataframe strategy. The setting of image parameters, large quantities parameters, and class modes ensures conformity with the specified data structures. The findings confirm the recognition of 695 validation samples and 2,780 samples for training, all of which are citizens of two distinct groups. The generators are designed for ensuing utilize in preparing machine learning models.
- Definition of model architecture: this model characterizes a convolutional neural network to organize (CNN) demonstrate utilizing the EfficientNetB0 design pre-trained on ImageNet. The amount of weight of the initial model remains constant and then more layers are added, such as the global average pooling, dense, dropout, and output layers, to create the entire system. The optimizer developed by Adam and categorised loss of cross-entropy are used to put collectively the model. The "fit" technique is employed to start the process of training, and data is supplied by means of generators of data that have already been set up. The preparing history is put away within the `history` variable for advance examination.
- Model evaluation: the description of the assesses the prepared model's execution of the approval information utilizing the evaluation strategy with the required information generator. The outcomes of the assessment are shown, including loss of validation and accuracy. The loss of validation is 0.4873, the accuracy is 81.43%, precision isa 51.16, recall is 50.11and F-score is 46.37as shown in the output. The model that has been trained is subsequently saved to a file through the utilisation of the save function.

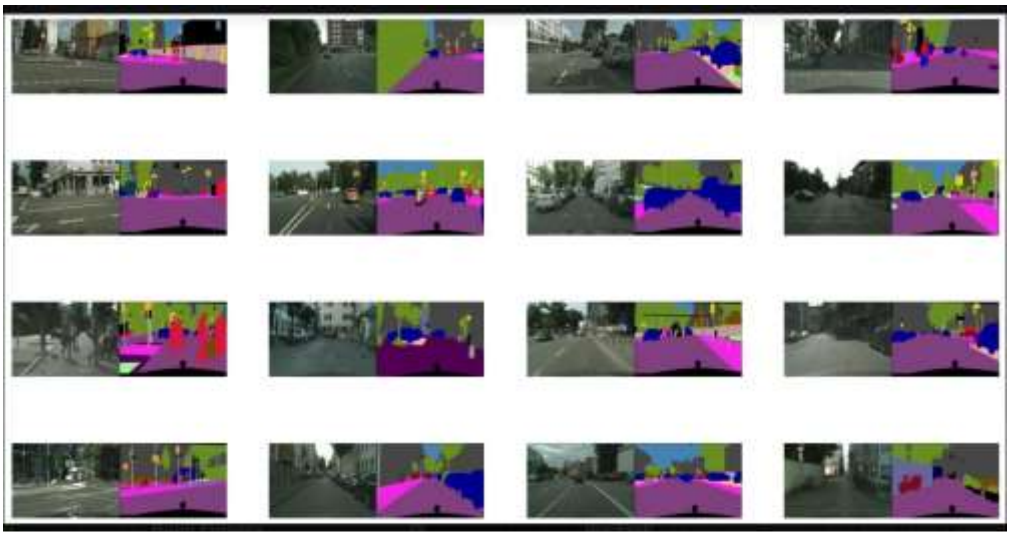

Figure 3. Description of images

Figure 4 blue color text can see calculated metrics values and in the x-axis of the confusion matrix chart indicates "predicted labels," whereas the y-axis indicates "true labels." and then diagonal blue and yellow boxes shows the accurate predicted count, while the remaining boxes have the inaccurate count. Table 2 it shows our methodology shows significant advantages over other established techniques in the comparative study of object detection, especially in accuracy, outperforming YOLOv5, the faster-RCNN with ResNet50, and depth wise deep convolution neural network (DDCNN).

Figure 5 In the graph, the x-axis represents different models or methods, while the y-axis represents the performance metrics. It has the highest accuracy of every model for this list. It does, however, exhibit reduced recall and precision when compared to some other models, suggesting a trade-off among preventing fake positives and correctly identifying positive results.

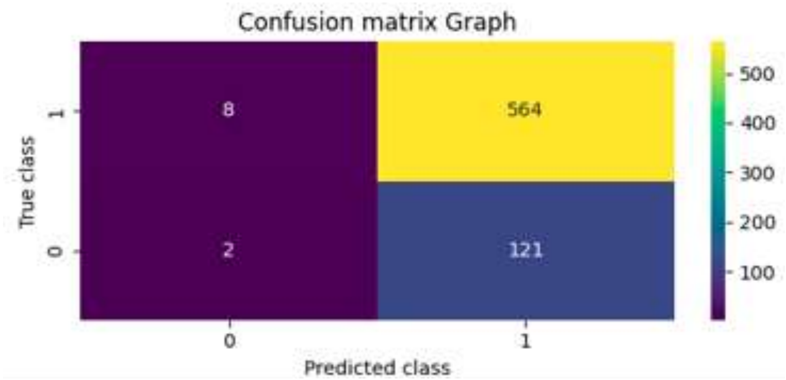

Figure 4. Confusion matrix with class labels for given datasets

Table 2. Evolution metrics of different method with ours

Method	Accuracy	Recall	Precision
YOLOv5 [19]	79.98	73.37	85.24
Faster-RCNN with ResNet50 [20]	78.41	52.71	36.35
DDCNN [21]	72.70	68.36	52.39
Ours	81.43	50.11	51.16

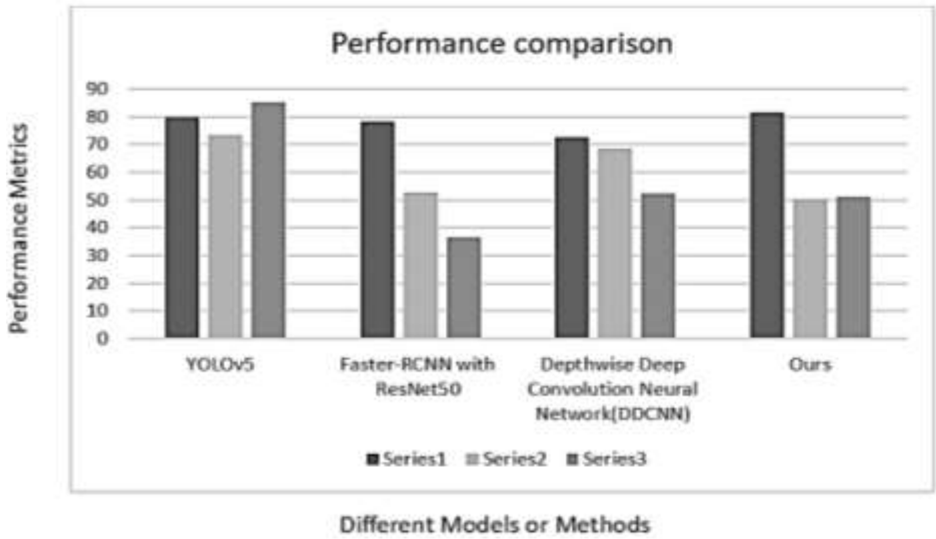

Figure 5. Performance comparison of classification with our model in accuracy, recall, and precision

3.1. Discussions

The displayed comes about typify the comprehensive improvement and assessment of a profound learning picture classification pipeline utilizing Google Colab. Through examining at the applied Python code, the developers observe three main things that show the significance it is in image processing and machine learning. In this example, the source code meticulously chooses and uses important Python libraries, with TensorFlow, Keras, and scikit-image playing key roles. The application of an array of image processing methods, XML parsing, and innovative deep learning algorithms such as DenseNet121 and MobileNetV2 demonstrates a sophisticated approach to creating a strong image classification system. The formal and systematic presentation of the code presents a careful strategy to extract optimal performance and ensures the reliability and efficiency of the implemented machine learning and image processing functions. This section of the implementation concentrates on the smooth integration of Google Drive with Google Colab [22], offering a workable solution for effective managing data in the cloud. The code creates a unified file access point by assigning the "extract_path" argument to a particular location within Google Drive, which expedites the data processing process. This section of the code concentrates on the smooth integration of Google Drive with Google Colab, offering a workable solution for effective cloud data management. The code creates a unified file access point by assigning the "extract_path" argument to a particular location within Google Drive, which expedites the information processing process. This interface facilitates easy sharing of code files and datasets across users, while also improving individual productivity and enabling effortless collaboration. Additionally, the method foresees future requirements by guaranteeing that acquired image data is easily accessible for further examination and training of models [23]. This combination is especially helpful in the Colab environment for managing big datasets and promoting collaborative. To manage the zipped image files found in the ZIP archive called IMAGE.zip. Utilizing the "extracted_path" result and "zip_file_path" communication, Google's Drive effectively finds a ZIP files [24]. By making it easier to create directories and extricate files, and enumerate folders inside the IMAGE path, this method streamlines the organization and processing of data [25].

4. CONCLUSION

In conclusion, it may be said that the research paper aims to change urban question recognizable proof by making a Python-based program that utilizes progressed calculations. This research work differs from the file model in that it only focuses on capturing infrared thermal images at night. Instead, it solves the difficult task of identifying objects in urban environments. The objectives include advertising a productive approach for identifying objects in urban situations, presenting a one-of-a-kind program based on Python and carrying out broad tests to exhibit predominant execution in different urban scenarios. The research work is presented in the framework of other studies, the main one being "the performance of important tests to show the effectiveness of the suggested model in comparison using the new models in globally." Our research has significantly advanced urban object recognition by integrating bidirectional GAN and MVAM. Our findings not only show increased recognition but also indicate a fundamental change in technological approaches. The success of the bidirectional GAN and MVAM integration opens promising avenues for future developments in urban studies and technology applications. This study establishes the foundation for a new era in urban object recognition, encouraging further exploration and refinement of these advanced methodologies to address the evolving challenges of urban environments. Future scope: Our research shows that combining bidirectional GAN and MVAM is a strong method for recognizing objects in urban areas. Future studies can explore ways to improve this approach, such as adjusting the settings of bidirectional GAN for different urban scenarios and trying different variations of the MVAM model.

REFERENCES

[1] C. Wang, *et al.* "Evolutionary generative adversarial networks," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 6, pp. 921-934, 2019, doi: 10.1109/TEVC.2019.2895748.

[2] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3943-3956, 2019, doi: 10.1109/TCSVT.2019.2920407.

[3] K. Liu *et al.* "FISS GAN: s generative adversarial network for foggy image semantic segmentation," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 8, pp. 1428-1439, 2021, doi: 10.1109/JAS.2021.1004057.

[4] A. Abdollahi *et al.* "Deep learning approaches applied to remote sensing datasets for road extraction: a state-of-the-art review," *Remote Sensing*, vol. 12, no. 9, pp. 1444, 2020, doi: 10.3390/rs12091444.

[5] P. Sun *et al.* "R4 Det: refined single-stage detector with feature recursion and refinement for rotating object detection in aerial images," *Image and Vision Computing*, vol. 103, 104036, 2020, doi: 10.1016/j.imavis.2020.104036.

[6] Z. Tian *et al.* "Object detection in optical remote sensing images by integrating object-to-object relationships," *Remote Sensing Letters*, vol. 11, no. 5, pp. 416-425, 2020, doi: 10.1080/2150704X.2020.1722330.

[7] X. Hua *et al.* "A fast self-attention cascaded network for object detection in large scene remote sensing images," *Applied Soft Computing*, vol. 94, p. 106495, 2020, doi: 10.1016/j.asoc.2020.106495.

[8] L. Zhang and J. Ma, "Salient object detection based on progressively supervised learning for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9682-9696, 2021, doi: 10.1109/TGRS.2020.3045708.

[9] X. Sun *et al.* "PBNet: part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 50-65, 2021, doi: 10.1016/j.isprsjprs.2020.12.015.

[10] E. Rani, "LittleYOLO-SPP: a delicate real-time vehicle detection algorithm," *Optik*, vol. 225, p. 165818, 2021, doi: 10.1016/j.ijleo.2020.165818.

[11] F. Fengxin *et al.* "AlTwo: vehicle recognition in foggy weather based on two-step recognition algorithm," *Advances in Neural Networks–ISNN 2020: 17th International Symposium on Neural Networks, ISNN 2020, Cairo, Egypt, December 4–6, 2020, Proceedings 17*. Springer International Publishing, 2020.

[12] Q. Huang, Z. Cai, and T. Lan, "A single neural network for mixed style license plate detection and recognition," *IEEE Access*, vol. 9, pp. 21777-21785, 2021, doi: 10.1109/ACCESS.2021.3055243.

[13] U. Naseem *et al.* "A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1-35, 2021, doi: 10.1145/3434237.

[14] P. Bhatt *et al*, "Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions," *Brain Informatics*, vol. 10, no. 1, p. 18, 2023, doi: 10.1186/s40708-023-00196-6.

[15] M. Chafiq, A. Chaouiki, and Y. G. Ko, "Recent advances in multifunctional reticular framework nanoparticles: a paradigm shift in materials science road to a structured future," *Nano-Micro Letters*, vol. 15, no. 1, p. 213, 2023, doi: 10.1007/s40820-023-01180-9.

[16] F. Kleber *et al.* "Exploration of the vienna city library poster collection using computer vision approaches," *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, IEEE, 2022, doi: 10.1109/ICECCME55909.2022.9988611.

[17] J. Fan *et al.* "A multi-class partial hinge loss for partial label learning," *Applied Intelligence*, vol. 53, no. 23, pp. 28333-28348, 2023, doi: 10.1007/s10489-023-04954-1.

[18] J. Lago *et al.* "Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark," *Applied Energy*, vol. 293, p. 1169832021, doi: 10.1016/j.apenergy.2021.116983.

[19] H. T. Vo and L. D. Quach, "advanced night time object detection in driver-assistance systems using thermal vision and YOLOv5," *International Journal of Advanced Computer Science and Applications*, vol.14, no. 6, 2023, doi:10.14569/IJACSA.2023.01406124.

[20] Y. C. Poh and S. S. Teoh. "Performance evaluation of optic disc detection using faster rcnn with alexnet, resnet50 and vgg19 convolutional neural networks," *Proceedings of the 11th International Conference on Robotics, Vision, Signal Processing and Power Applications: Enhancing Research and Innovation through the Fourth Industrial Revolution*. Singapore: Springer Singapore, 2022, pp 753–758

[21] H. Patel *et al.* "Depthwise convolution for compact object detector in nighttime images," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[22] S. Aoki, T. Higuchi, and O. Altintas, "Cooperative perception with deep reinforcement learning for connected vehicles," *2020 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2020, doi: 10.1109/IV47402.2020.9304570.

[23] B. Maheswari and S. R. Reeja, "Thermal infrared image semantic segmentation for night-time driving scenes based on deep learning," *Multimedia Tools and Applications*, pp. 1-26, 2023, doi: 10.1007/s11042-023-15882-0.

[24] H. Wu *et al.* "A weakly supervised learning method for cell detection and tracking using incomplete initial annotations" *International Journal of Molecular Sciences*, vol. 24, no. 22, p. 16028, 2023, doi: 10.3390/ijms242216028.

[25] A. Abdulmunem, N. Al-Shakarchy, and M. Safoq, "Deep learning based masked face recognition in the era of the COVID-19 pandemic," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 1550-1559, 2023, doi: 10.11591/ijece.v13i2.pp1550-1559, doi: 10.11591/ijece.v13i2.pp1550-1559.

BIOGRAPHIES OF AUTHORS

Maheswari Bandi Earned her undergraduate degree from Lakireddy Bali Reddy College of Engineering, Vijayawada, JNTUK, and her Master's degree in Computer Science and Engineering from Vikas College of Engineering and Technology, Vijayawada, JNTUK, Andhra Pradesh. She is having seven years of experience in teaching and presently she is pursuing a Ph.D. degree at the School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Andra Pradesh (VIT-AP University). Her current research interests are in the area of deep learning, digital image processing, and machine learning. She can be contacted at email: mahesawari.21phd7152@vitap.ac.in.

Reeja Sundaran Rajakumari Professor and Head, Department of Artificial Intelligence and Machine Learning, Vellore Institute of Technology, Andra Pradesh (VIT-AP University). Her research area is artificial intelligence, high-performance computing, digital image processing, and computer vision. She earned her Ph.D. in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Govt. of Karnataka for her thesis Real-Time Video Denoising. She has 18 years of teaching experience in various engineering colleges as well as Universities with 12 years of research experience in the concerned field. She has published in many SCOPUS journals, national and international journals, book chapter publications, international conferences, and national conferences. She has received many awards- Meritorious alumni award (BTech), best paper award, best student award during M.Tech, and best outgoing student award during M. Tech and School level Kalathilakam in 10th standard. She has also a Programme committee member, reviewer, and session chair for the various National and international conference and seminars. She can be contacted at email: reeja.sr@vitap.ac.in.