

Hybrid Deep Neural Network-Based Alert System for Wild Animal Activity Identification

1. M . Y. Durga Tejaswi
Assistant Professor, Department of CSE
Rajeev Gandhi Memorial College of Engineering
& Technology, Nandyal, India
Email: yagnaputejaswi@gmail.com

2. Mr. .C.K. Venkata Tharun
MCA Scholar, Department of MCA
Rajeev Gandhi Memorial College of Engineering
& Technology, Nandyal, India
Email: tharunack@gmail.com

Abstract: This research addresses the pressing problem of the appearance of animal attacks in rural areas and forestry staff by developing an efficient surveillance system. The suggested Hybrid Visual Geometry Group (VGG)-19 with Bidirectional Long Short-Term Memory (Bi-LSTM) network aims at recognizing animal species, monitoring their activity, and providing timely locational information to send alerts in forest areas. Using the VGG-19 to extract features and Bi-LSTM to learn sequences, the model achieves the accuracy of animal recognition and movement patterns, as opposed to other traditional surveillance methods. As well, an ensemble approach is applied, which combines the forecast of different independent models to enhance resilience and accuracy. In addition to that, the performance is greatly increased with the use of methods like CNN supplemented with BiGRU, reaching the accuracy of 100 percent. The implementation provides an easy-to-use front end coded in Flask framework with user testing being possible with authentication. This paper is a feasible plan to minimize the risks of animal attacks, where advanced deep learning techniques and user-friendly design are used to implement effective safety measures in rural and forestry areas.

Index Terms – *Animal detection, VGG-Net, Bi-LSTM, convolutional neural network, activity recognition, video surveillance, wild animal monitoring, alert system.*

I. INTRODUCTION

The recent years have seen the rising rates of human-wildlife interactions which have revealed the dire need of new approaches that can reduce the conflict rates and enhance the safety levels of humans and wildlife. The urbanization and agricultural development drove by habitat invasion has increasingly brought people into the close contact with animals leading to the rise of conflicts and threats [1]. These interactions have proven to be very challenging to conservationists, policymakers, and local communities, and as a result, the creation of effective early warning systems is necessary to minimize risks and to provide timely solutions.

Wild Animal Activity Alerting System (WAAAS) is one of the possible solutions to the pressing problems of the interaction between humans and wildlife. With the help of advanced Machine Learning (ML) and DL frameworks, WAAAS aims at detecting and pointing out wild animal behavior near the human settlements, which provides the necessary data used in preemptive management strategies [2]. WAAAS draws on information through multiple channels such as photos, movement sensors and sound clips to determine

trends that indicate the existence and activity of the wildlife to prevent inaccurate and delayed alerts [3].

The need to have such a system is brought out by the fact that there is an escalation of conflicts between humans and wildlife due to habitat fragmentation and resource competition. These wars can be expressed in various ways such as raiding of herbivores, livestock predation, and infrequent human attacks. These actions put the lives of humans and livelihoods in danger, destroy the lives of wild animals and hamper conservation. This has led to growing acceptance of the need to employ active measures to mitigate the conflicts and ensure peaceful co-existence between people and wildlife.

The development of WAAAS is an important step of enhancing the safety level, minimizing the economic losses, and promoting conservation objectives. WAAAS is a ML and DL based method that can analyze wild animal action signals with an unprecedented precision and efficiency [6]. This introduction explores the intricate nature of human-wildlife interactions and the importance of early warning systems in addressing those questions, as well as, explains the objectives and scope of

WAAAS in the reduction of conflicts and encouragement of coexistence.

A Growing Menace Human-wildlife interaction has been a major international concern which has grown due to habitat fragmentation, climate change, and rising levels of human population. The human-wildlife conflicts have been increasing in the different ecosystems due to the decline in natural environments as human activities encroach into the habitats and the degradation of their natural habitats. Agricultural losses and predation of livestock are experienced in rural communities, whereas urban communities are experiencing wildlife intrusion and destruction of properties which have varied ramifications.

Crop pesting especially by the elephants, deer and wild boars is a significant risk to the agricultural livelihoods leading to enormous economic losses and heightening food poverty among the vulnerable groups [10]. Similarly, livestock carnivores (wolves, lions, and bears) cause conflict between pastoralists and conservationists, often leading to retaliatory killing and escalation of conservation problems [11]. Moreover, the periodical attacks on humans by big predators do not only bring about fear and insecurity, but also increase the tension between the locals and the wildlife authorities [12].

There are also complications related to alleviating human-wildlife conflicts due to the dynamism of wildlife behavior as well as the diversity of species. All species have their specific feeding behaviors, territorial activities, and response to disturbance caused by humans and so each one needs unique management interventions [13]. The temporal and spatial heterogeneity of the anthropogenic-wildlife interactions underscores the need to have real-time monitoring and the early warning systems to facilitate proactive responses [14].

II. RELATED WORK

The increased attention to the study of the concept of scene has led to the emergence of the critical work on developing the strategies aimed at understanding complex visual images. Aarthi and Chitrakala [1] provide a comprehensive overview of the methodologies of the scene interpretation focused on the diversity of strategies applied to the sphere. Their study shows the importance of understanding scenes in various applications including robots, autonomous navigation and surveillance.

CSTs give a strong description of both the structure and the hierarchy of the regions in an image. Ahuja and Todorovic [2] introduce the concept of CSTs which make it possible to represent spatial arrangement and hierarchy jointly, thereby

improving the picture segmentation and the efficiency of the object recognition. This approach has been applied in other areas including medical imaging, remote sensing, as well as analysis of scenes.

There has been a wide application to ML algorithms in predictive modeling processes, including disease prediction. The study by Assegie et al. [3] is an empirical study on the use of ML in predicting heart illness, and it evaluates the effectiveness of a variety of classifiers on a heart disease dataset. The results of their research shed some light on how some of the algorithms can be used to predict the risk of heart disease and this information can be used in clinical decision-making as well as risk stratification.

DL has revolutionized the field of computer vision and made it possible to achieve impressive advancements in the understanding of objects and their detection. Banupriya et al. [4] focus on applying DL techniques to animal recognition and present the high efficiency of CNNs to identify and classify animals in images. The literature reveals how the DL-based methodology can be used in the wildlife surveillance and conservation programs.

Objectness estimate is a critical element of object detection tasks, which helps to identify areas that are likely to have objects with the interest. According to Cheng et al. [5], Binarized Normed Gradients (BING) is an objectness estimator that achieves real-time results of 300 frames per second (fps). BING employs normalized gradients and binarized features in offering a lightweight and efficient object detection solution in pictures.

Another example of the state of the art approach to object detection is R-FCNs which involves the localization of objects and classification through region-based techniques. R-FCN offered by Dai et al. [6] is a hybrid of FCNs and RPNs that can be used to identify objects on photos efficiently. Their study demonstrates the performance of R-FCN in achieving competitive performance on benchmark data.

The need of change detection in many applications, including environmental monitoring, urban planning, and surveillance is critical. De Gregorio and Giordano [7] suggest a framework of change detection based on weightless neural networks that have adaptation to changing environments and resilience to noise. Their approach demonstrates promising results in discovering the changes in remote sensing images, which helps in land cover mapping and catastrophe management.

Sign language recognition is a daunting task, which is due to complex and dynamic nature of sign

motions. Natarajan et al. [8] introduce an overall DL system to identify sign language, translate and create videos. Their study employs the deep neural networks to identify information contained in the sign language videos, so that sign movements are correctly detected and translated and can therefore offer a feasible means through which communication can be improved among the hearing handicapped.

III. MATERIALS AND METHODS

The developed project is based on the design and evaluation of a hybrid VGG-19 + Bi-LSTM network that would be used to recognize the presence of animals in forested regions. This method utilizes the combination of feature extraction and sequence learning to improve the accuracy and enable real-time monitoring to improve the safety by providing timely warnings. Moreover, the project expansion will imply the introduction of CNN+GRU model, which will operate with the help of one of the Bidirectional GRU layers and applying the CNN algorithm to their results to enhance precision. GRU was selected because it is more effective at picture feature optimization compared to LSTM. Also, a Flask application based on SQLite has been developed to simplify user registration and authentication, therefore, enabling the user to test the features of the system and enhance usability. The additions aim at providing a holistic and effective solution towards addressing the problems of detecting animal activity in forest ecosystems.

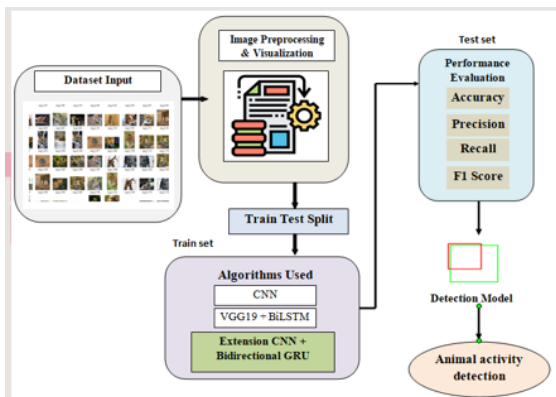


Fig.1 Proposed Architecture

The system design starts with the input data which is the photographic data collected in the forested areas. Preprocesses are done on the images to generate normalization and augmentation before the division of the images into training and testing sets. The training stage includes the use of three specific algorithms: CNN, VGG19-BiLSTM, and CNN+Bidirectional GRU that are specifically trained to detect animal activity. The CNN model focuses on convolutional layers in extracting

features, whereas the VGG19-BiLSTM integrates convolutional layers with Bi-LSTM in sequence learning. CNN+Bidirectional GRU model is CNN with a Bidirectional GRU layer to boost the accuracy. The models are then evaluated with an independent test set after training to test their effectiveness in detecting animal behaviors. The real-time monitoring-enabled detection model processes the received photos and sends timely alerts in case of animal movement in the vicinity, thereby enhancing security in the forest cover.

A) Dataset Collection:

The data gathering method involved getting photos of four different benchmark datasets namely: the camera trap dataset [48], the wild animal dataset [49], the hoofed animal dataset [50] and the CDnet dataset [51]. These data sets have very diverse habitats and wildlife species so the training and testing of the proposed model can be broadly covered. The camera trap data captures the action of wildlife in the wild whereas the wild animal data captures a broader set of species and actions. Of particular interest to the hoofed animal dataset is the focus on species-specific characteristics of the training data, which focuses on the mammals with hooves. The CDnet dataset makes the task more diverse with annotated video sequences that should be evaluated. The data gathering method involved getting photos of four different benchmark datasets namely: the camera trap dataset [48], the wild animal dataset [49], the hoofed animal dataset [50] and the CDnet dataset [51].



Fig.2 Dataset Collection

These datasets cover numerous habitats and wildlife species, which provides comprehensive coverage of the proposed model training and testing. The camera trap dataset reports on the activity of wildlife in the wild, but the wild animal dataset shows a more diverse species and behavior. The hoofed animal dataset particularly focuses on hoofed mammals and therefore increasing the training data with species-specific characteristics. The CDnet data adds some variability by annotated video sequences to be evaluated. Through the combination of multiple datasets, the model is provided with a complex and

diverse training corpus, which allows the successful recognition of animal activity in the wild in changing environmental conditions and species groups.

B) Processing:

The preprocessing stage of the data is a sequence of procedures that are necessary to ensure that deep neural networks are performing phenomenally when it comes to identifying the behavior of wild animals. To begin with, the images are normalized to normalize their pixel values and this enhances uniformity across the dataset and convergence in training. The photos are then mixed up to randomize their order and hence reduce bias and boost the generalization ability of the model. Moreover, the dataset can undergo augmentation processes (rotation, flipping, or cropping) to increase both its variety and resiliency. There are also labels to the photographs to show the activity of the wild animals or not. The pre-processing step prepares the work of the effective training and evaluation of hybrid deep neural networks through the extensive normalization, shuffling, and marking of the data set. This ensures that the models are able to accurately determine and classify wild animal behavior in varying environmental conditions and hence give timely alarm messages to enhance safety in wildlife grounds.

Visualization: Visualization with the help of Seaborn and Matplotlib is necessary to extract insights out of the data and evaluate the efficiency of the model. These libraries are rich in producing useful plot and chart generating tools such as histograms, scatter plots and heatmaps. High level interface Seaborn is built on Matplotlib and allows users to build visually appealing visualization with no effort or coding. These visualizations help to understand how the data is distributed, find the outliers, and see the trends which can impact the model training and performance. Moreover, the presentation of model parameters such as accuracy, loss, and confusion matrices will make a cumulative assessment of the model performance and effectiveness. Researchers can effectively present the results and address the hypotheses, as well as make informed decisions along the ML pipeline, using Seaborn and Matplotlib, thereby increasing the strength and reliability of the research and the models created.

Feature Extraction: An important process used in identifying wild animals behavior using hybrid deep neural networks is feature extraction. In this regard, feature extraction is the study of the input photos to identify meaningful patterns and attributes that would distinguish different types of animal behavior. CNNs are also commonly used in feature-

extraction since they have the capability of learning hierarchical findings on raw pixel data in an autonomous manner. The hybrid model proposed in the present paper employs a mixture of CNN and RNNs), with the use of LSTM and Gated Recurrent Units (GRUs) in the extraction of features. CNN element obtains spatial details of the input images, whereas, the RNN element obtains motion and context details of the temporal sequences. Combining these two types of neural networks makes the hybrid model an effective extractor of geographical and temporal information, thus, making it possible to accurately detect the presence of wildlife and trigger an alarm to the relevant authorities or citizen.

Training and testing: The creation of a robust system which is used in generation of alert messages based on the detection of wild animal behavior using hybrid deep neural networks requires testing and training as part of the process. The hybrid model is introduced during training with a labeled dataset with photos of various examples of wild animal behavior. Through the use of iterative optimisation of model parameters i.e. the use of backpropagation, the model learns the process of identifying the relevant information and producing accurate predictions. Training involves evaluating the effectiveness of the model on a separate test of the data to ensure the effectiveness of the model to new circumstances.

Testing on the other hand evaluates the effectiveness of the trained model on a different set of data, which is a realistic scenario. Precision in the identification of activities of wild animals and generation of alert notices: the proficiency of the model in terms of its ability to identify wild animal activities accurately and precisely is measured by the following: accuracy, precision, recall, and F1, etc. Testing also helps the researchers to evaluate the effectiveness of the model, identify possible gaps, and refine the architecture or training process accordingly. Finally, the development of a reliable system aimed at effective alleviation of the risks concerning wild animals interactions requires thorough testing and training.

C) Algorithms:

Existing CNN: Convolutional Neural Network (CNN) is a DL model that is specifically designed to be used in image recognition. The identification of the activities of wild animals in this study is carried out using an existing CNN[16]. CNNs have convolutional layers which extract features of the input images, followed by pooling layers to reduce dimensions and fully connected layers to make the classification. The study uses CNN to decode image data in the raw form, where relevant features that

depict the presence of wildlife are extracted. With the capabilities of the CNN [16] and hierarchical feature extraction, the model can effectively generate the patterns of the images and classify them to generate an alert message in response to the movement of the animal(s) that is noted.

Propose VGG19 + BI-LSTM: The proposed VGG19 + BI-LSTM is a hybrid model that integrates VGG19 CNN model with the Bidirectional Long Short-Term Memory (BI-LSTM) recurrent neural networks. VGG19 is used to obtain hierarchical features of input photos and BI-LSTM is used to obtain time dependence features by analyzing sequential data. This is a hybrid model which is used to detect wild animal activities in the project. VGG19 obtains spatial characteristics of photos, and BI-LSTM obtains time series of feature vectors where it detects patterns that can signal a change in the behavior of animals over time. The VGG19 + BI-LSTM model enhances the accuracy of the detection of activities of wild animals, combining spatial and time data, which allows sending an alert as quickly as possible.

CNN + GRU: CNN + GRU model combines the use of CNNs and GRUs. CNNs learn the spatial characteristics of images, and GRUs events in sequential data to learn the temporal characteristics of them. This is a hybrid model which is used to detect wild animal activities in the project. CNN component extracts spatial features of input photos, and GRU component processes the time series of the animal feature vectors to extract patterns that could be evidence of a change in behavior. The CNN + GRU model enhances the accuracy of detecting the activity of wild animals by adding spatial and temporal information, thereby allowing alert notifications to be issued in a timely manner.

This can be easily achieved by a simple formula:

Dimension of image = (n, n)

Dimension of filter = (f, f)

Dimension of output will be ((n-f+1), (n-f+1)) (1)

At this point, you should have a full and in-depth observation on how a convolutional layer works. We shall now turn to the next part of CNN structure.

GRU:

The roles of a GRU can be expressed as the following equations:

- **Update Gate:** $z_t = \sigma(W_z * [h_{t-1}, x_t])$

- **Reset Gate:** $r_t = \sigma(W_r * [h_{t-1}, x_t])$
- **Candidate Hidden State:** $\tilde{h}_t = \tanh(W * [r_t \odot h_{t-1}, x_t])$
- **Final Hidden State:** $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$

Here, σ represents the [sigmoid function](#), \tanh is the hyperbolic tangent function, W_z , W_r , and W are parameter matrices, h_{t-1} is the previous hidden state, x_t is the current input, \odot represents element-wise multiplication, and h_t is the current hidden state.

IV. RESULTS AND DISCUSSION

Accuracy: Accuracy of a test is the ability to distinguish between patients and healthy people. In order to determine the accuracy of the tests; calculate the ratio between false positives and false negatives of all the tests done. This may be mathematically expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Precision: Precision measures the percentage of the cases or samples that are correctly identified. The formula of determining precision is:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

Recall: ML recall is used to determine a model that can identify every relevant example of a class. It shows the effectiveness of a model in capturing the number of instances of a class by comparing correctly observed positives to the positives.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-Score: The F1 score is used to determine the accuracy of a ML model. Combining model recall and precision. The accuracy measure is used to gauge how many predictions are true by a model of the data.

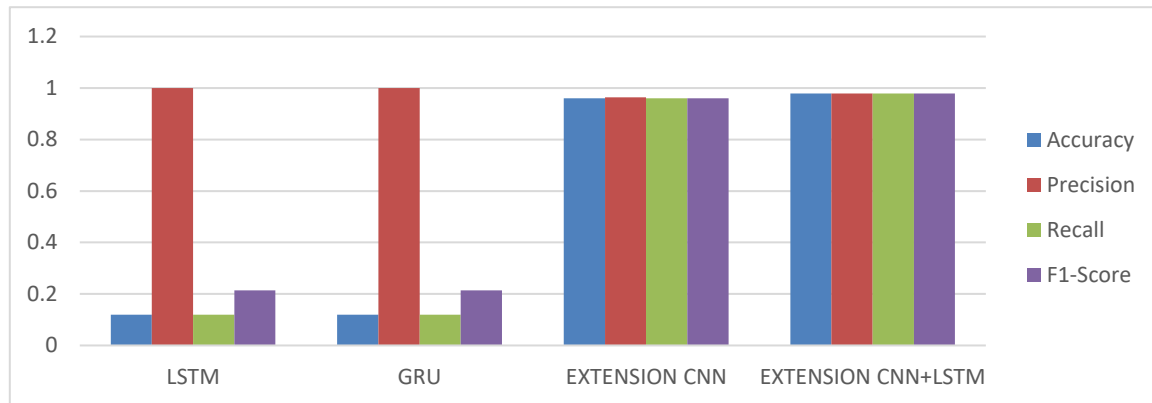
$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} * 100 \quad (5)$$

Table (1) evaluates the performance measurements such as accuracy, precision, recall and F1-score of each algorithm. CNN+GRU extension outperforms all other algorithms on a regular basis. The tables will offer a comparative analysis of the measures of the alternative methods.

Table.1 Performance Evaluation Table

ML Model	Accuracy	Precision	Recall	F1-Score
LSTM	0.120	1.000	0.120	0.214
GRU	0.120	1.000	0.120	0.214
EXTENSION CNN	0.960	0.964	0.960	0.960
EXTENSION CNN+LSTM	0.979	0.979	0.979	0.979

Graph.1 Comparison Graphs



Graph (1) represents accuracy in blue, precision in red, recall in green and F1-Score in purple. In comparison to the previous ones, the extension CNN +GRU is more performing in all measures, achieving the greatest values. These findings are graphically illustrated in the above graphs.

V. CONCLUSION

Finally, the development and integration of CNN, VGG19 + BI-LSTM, and CNN-GRU algorithms have demonstrated excellent effectiveness in the classification of wild animal behaviors. Minimized computational expenses due to streamlined procedures are guaranteed to make forest monitoring programs sustainable. A hybrid CNN-GRU model, where the benefits of both CNN and GRU are used, has proven stronger detection capabilities, which are even more effective than each single model. The introduction of a user-friendly Flask front-end and a SQLite database enhance the usability of the system and support the effective input of data and displaying the animal detection results. This project, in the end, brings a lot of benefit to the rural populations, forestry employees and conservationists through the provision of an efficient, effective, and precise tool on the monitoring of wildlife activities. This project boosts security and preservation in the forested regions and this is beneficial both to human beings and wildlife and clear coexistence in the natural environment.

The alert message production system that is based on the detection of wild animal behavior with the

help of hybrid deep neural networks has one of the feature scopes that presuppose a lot of necessary elements. The system is also put in a way that it accurately recognizes and classifies different animal behaviors and patterns in the natural environments, including movements, interactions, and weirdnesses. The technology is to investigate temporal patterns of animal behavior in order to detect trends that can be viewed as possible dangers or disruptions. Also, the capabilities of the system include optimization of computational costs and sustainability in forest monitoring programs. It also includes the delivery of a user-friendly interface that offers an easy data entry and presentation of the animal detection findings. Its feature scope will be to advance safety and conservation in wooded regions by giving timely notifications to the relevant stakeholders, e.g., rural population, forestry laborers, and conservationists, which promotes peaceful coexistence between human beings and wild animals.

REFERENCES

- [1] S. Aarthi and S. Chitrakala, "Scene understanding—A survey," in Proc. Int. Conf. Comput., Commun. Signal Process. (ICCCSP), Jan. 2017, pp. 1–4.
- [2] N. Ahuja and S. Todorovic, "Connected segmentation tree—A joint representation of region layout and hierarchy," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1–8.
- [3] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari, "An empirical study on machine learning algorithms for heart disease prediction," IAES Int. J. Artif. Intell. (IJ-AI), vol. 11, no. 3, p. 1066, Sep. 2022.
- [4] N. Banupriya, S. Saranya, R. Swaminathan, S. Harikumar, and S. Palanisamy, "Animal detection using deep learning algorithm," J. Crit. Rev., vol. 7, no. 1, pp. 434–439, 2020.

- [5] M. Cheng, Z. Zhang, W. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 3286–3293.
- [6] J. Dai, Y. Li, K. He, and J. Sun, "RFCN: Object detection via region-based fully convolutional networks," 2016, arXiv:1605.06409.
- [7] M. De Gregorio and M. Giordano, "Change detection with weightless neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2014, pp. 409–413.
- [8] B. Natarajan, E. Rajalakshmi, R. Elakkiya, K. Kotecha, A. Abraham, L. A. Gabralla, and V. Subramaniaswamy, "Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation," IEEE Access, vol. 10, pp. 104358–104374, 2022.
- [9] W. Dong, P. Roy, C. Peng, and V. Isler, "Ellipse R-CNN: Learning to infer elliptical object from clustering and occlusion," IEEE Trans. Image Process., vol. 30, pp. 2193–2206, 2021.
- [10] R. Elakkiya, P. Vijayakumar, and M. Karuppiah, "COVID_SCREENNET: COVID-19 screening in chest radiography images using deep transfer stacking," Inf. Syst. Frontiers, vol. 23, pp. 1369–1383, Mar. 2021.
- [11] R. Elakkiya, V. Subramaniaswamy, V. Vijayakumar, and A. Mahanti, "Cervical cancer diagnostics healthcare system using hybrid object detection adversarial networks," IEEE J. Biomed. Health Informat., vol. 26, no. 4, pp. 1464–1471, Apr. 2022.
- [12] R. Elakkiya, K. S. S. Teja, L. Jegatha Deborah, C. Bisogni, and C. Medaglia, "Imaging based cervical cancer diagnostics using small object detection—Generative adversarial networks," Multimedia Tools Appl., vol. 81, pp. 1–17, Feb. 2022.
- [13] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in Computer Vision—ECCV. Dublin, Ireland: Springer, Jun. 2000, pp. 751–767.
- [14] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2155–2162.
- [15] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in Proc. 13th Scand. Conf. (SCIA). Halmstad, Sweden: Springer, Jul. 2003, pp. 363–370.
- [16] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.
- [17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.
- [18] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Change detection: Net: A new change detection benchmark dataset," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 1–8.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
- [20] J. Imran and B. Raman, "Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition," J. Ambient Intell. Humanized Comput., vol. 11, no. 1, pp. 189–208, Jan. 2020.
- [21] F. Kahl, R. Hartley, and V. Hilsenstein, "Novelty detection in image sequences with dynamic background," in Statistical Methods in Video Processing, Prague, Czech Republic: Springer, May 2004, pp. 117–128.
- [22] T. Liang, H. Bao, W. Pan, and F. Pan, "Traffic sign detection via improved sparse R-CNN for autonomous vehicles," J. Adv. Transp., vol. 2022, pp. 1–16, Mar. 2022.
- [23] T. Liang, H. Bao, W. Pan, X. Fan, and H. Li, "DetectFormer: Category assisted transformer for traffic scene object detection," Sensors, vol. 22, no. 13, p. 4833, Jun. 2022.
- [24] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 478–487.
- [25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Computer Vision—ECCV. Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.
- [26] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 831–843, Aug. 2000.
- [27] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1717–1724.
- [28] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 2056–2063.
- [29] M. Monnet, P. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in Proc. 9th IEEE Int. Conf. Comput. Vis., 2003, pp. 1305–1312.
- [30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.