A Review of Convolutional Neural Network-Based Object Detection in UAV Thermal Infrared Videos for Human-Elephant Conflict Response

AHKK Pathmadewa

Doctoral Candidate, Faculty of Graduate Studies, General Sir Jhon Kotelawala Defence University, Sri Lanka

kkpathmadewa@yahoo.com

Abstract: *Human-Elephant Conflict (HEC) is a significant issue in regions where human settlements intersect with elephant habitats, posing a threat to lives and livelihoods. Unmanned Aerial Vehicles (UAVs) equipped with thermal infrared (TIR) cameras offer a promising solution for real-time monitoring and management in HEC situations. This review paper comprehensively examines the integration of Convolutional Neural Network (CNN) algorithms into UAV video streaming to enhance object detection in TIR videos. The paper begins with an overview of HEC mitigation strategies, highlighting the role of UAV surveillance and the need for accurate object detection algorithms to distinguish between Humans and Elephants in challenging environmental conditions. The background and related work section examines previous research on HEC mitigation and object detection techniques in UAV videos, specifically focusing on CNN-based approaches and challenges unique to TIR imagery. The paper then explores object detection algorithms tailored for TIR videos, detailing architectures like YOLO, SSD, and Faster R-CNN, and highlighting their strengths and limitations. Considerations of integrating CNN algorithms into UAV systems are discussed, addressing challenges such as computational efficiency and optimisation for TIR video streaming. The review also covers evaluation metrics and performance analysis, stressing the importance of precision, recall, F1 score, and Intersection over Union (IoU) in assessing algorithm effectiveness. Furthermore, the paper outlines future directions and challenges, including multi-sensor fusion and ethical considerations in deploying UAV technology for HEC management. In conclusion, the paper underscores the significance of CNN-based object detection in UAV TIR videos for emergency response in HEC situations, and the need for ongoing innovation and collaboration to mitigate human-elephant conflict effectively.*

Keywords: Object Detection, Convolutional Neural Networks, Human-Elephant Conflict, Unmanned Aerial Vehicle

Introduction

The continuous struggle for territory and resources between people and wild elephants is known as the "Human Elephant Conflict" (HEC) (Castaldo-Walsh, 2019). As populations grow, agricultural and other land activities encroach on natural habitats, and elephants frequently enter farmland in search of food, causing damage to crops and properties (George Wittemyer, 2008; Naughton-Treves, 2010; Nyhus, 2016). In response, farmers may use harmful methods to deter elephants, resulting in reactive actions from elephants and escalating conflict.

According to studies, Sri Lankan Forest Elephants, a distinct subspecies of the Asian elephant, prefer to hide in places that are difficult to see during the day, such as bushes and miniature forests. They are more active in open areas at night (Fernando et al., 2021). In the context of HEC, Unmanned Aerial Vehicles (UAVs) equipped with Infrared (IR) cameras can be used to detect elephants in human settlements or agricultural areas. Early detection allows authorities to respond quickly and prevent conflicts. The IR imagery assists in tracking elephant movements, understanding their behaviour, and deploying appropriate interventions.

UAVs equipped with thermal infrared (TIR) cameras provide a unique advantage for efforts to mitigate HEC. These drones enhance early detection and intervention by offering aerial surveillance capabilities, enabling prompt responses to potential conflict situations. Nevertheless, the success of using UAVs for monitoring depends heavily on the precision and speed of the object detection algorithms used to analyse the real-time TIR videos.

This review paper analyses the integration of Convolutional Neural Network (CNN) algorithms in TIR video streaming to detect objects, specifically focusing on elephants and humans. The paper comprehensively explores existing research and technologies and aims to understand the critical areas of state-of-the-art methodologies, challenges, and future directions.

Furthermore, the paper will examine the landscape of object detection techniques in UAV images and videos, ranging from conventional methods to cutting-edge CNN-based approaches. Special attention will be given to the nuances of TIR imagery, including its low resolution, noise, and variable environmental conditions, which pose unique challenges for accurate detection.

Background and Related Work

HEC mitigation has long been challenging in regions such as Sri Lanka, where humans' and elephants' habitats intersect. Traditional mitigation methods, including physical barriers, trenches, and deterrents such as noise-making devices, have proven effective to some extent (Hoare, 1999). However, these traditional methods often fail to provide comprehensive solutions, mainly when elephant movement patterns are unpredictable or when conflicts arise in remote areas. This highlights the need for more advanced and adaptable solutions, such as UAV-based surveillance.

Technological advancements have offered new avenues for addressing HEC in recent years, with UAVs emerging as a promising real-time monitoring and management tool (Gonzalez et al., 2016; Witczuk et al., 2018). UAVs equipped with TIR cameras have gained traction due to their ability to provide aerial surveillance capabilities, enabling early detection of elephant herds and human activities even in challenging terrain and lighting conditions.

The literature on HEC mitigation reflects a transition from reliance on traditional methods towards a more technology-driven approach, with UAV surveillance playing a pivotal role. Studies have demonstrated the effectiveness of UAVs in reducing response times to conflict incidents, enabling rapid deployment of intervention strategies, and minimising human-elephant confrontations.

a. Human Elephant Conflict

HEC in Sri Lanka is a complex and pressing issue with significant socioeconomic and environmental implications. The escalating conflict results from the intersection of human encroachment into traditional elephant habitats (Fernando, 2015; Fernando et al., 2021), habitat fragmentation (Köpke et al., 2021; Shaffer et al., 2019), and the elephants' natural migration patterns (Anuradha et al., 2019). As populations grow and human activities such as agriculture and settlements encroach on wildlife territories, the elephants are frequently forced to enter human settlements and farmland in search of food and water, increasing the likelihood of confrontations (Tiller et al., 2021).

Conflicts between humans and elephants often negatively affect communities and wildlife. On the one hand, communities suffer significant economic losses due to crop damage, infrastructure destruction, and even human casualties (Leimgruber et al., 2011; Santiapillai et al., 2010). Conversely, elephants face retaliatory killings, habitat loss, and injury due to human conflict (Billah et al., 2021; Das et al., 2014; Sitati et al., 2003). In this context, novel conflict resolution strategies are required to be introduced to mitigate the HEC.

In the emergency response to HEC, deploying UAVs equipped with TIR cameras appears to be a promising technological intervention (Akula et al., 2014; D'Acremont et al., 2019; Nasrabadi, 2019). The unique topography and vegetation of Sir Lanka challenge the traditional surveillance methods, making UAVs a practical and versatile tool for monitoring large and inaccessible areas. The TIR technology is advantageous because it detects thermal signatures, allowing elephants to be identified even in low-light conditions.

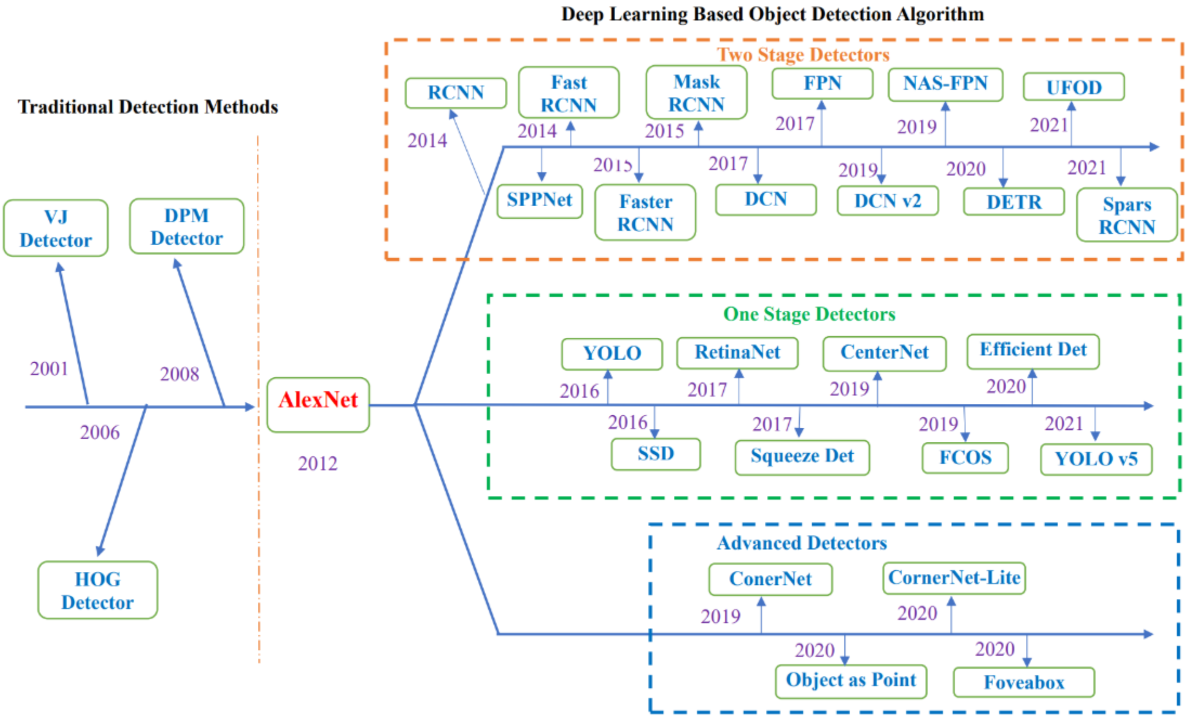
b. Traditional Object Detection Method

Due to CNN algorithm improvements, Object detection in images and videos has become a prominent field of study that has advanced significantly in recent years. Further, traditional methods were extensively used in object detection in images in early 2000, and they evolved with time in significant accuracy. In 2004, Viola–Jones (VJ) object detector(Viola & Jones, 2004) utilised a boosted cascade of simple features to achieve high accuracy and fast processing on face detection.

Further, Histograms of Oriented Gradients (HOG)(Dalal et al., 2005) features capture local gradient information to represent the appearance and shape of humans, and it achieves state-of-the-art results in human detection, demonstrating its effectiveness in various challenging scenarios. When it considers object detection using Deformable Part Models (DPM)(Felzenszwalb et al., 2008) in a cascade framework, it introduces a more flexible model incorporating parts with varying spatial relationships, enhancing the representation of complex objects. Since the traditional deep learning object detection method is reaching its upper limits, the requirement arises for a new system to identify the object more accurately.

c. Deep Learning-Based Object Detection Methods

In 2012, Krizhevsky, Sutskever and Hinton developed a ground-breaking image classification approach using deep CNNs. The authors train a large-scale CNN architecture called AlexNet on the ImageNet dataset, significantly improving classification accuracy. They introduced innovations such as Rectified Linear Units (ReLU), data augmentation, dropout regularisation, and GPU acceleration (Krizhevsky et al., 2012).



Source: Mittal et al., 2020; Wu et al., 2020

Figure 1: CNN-based object detection methods

The study demonstrates the power of deep CNNs in learning hierarchical features directly from raw pixels, enabling advanced performance on the challenging ImageNet dataset. The paper's contributions have revolutionised computer vision and deep learning, paving the way for subsequent advancements in image recognition tasks. After more than a decade, CNN-based object detection algorithms have become more advanced compared to previous studies. Considering the development of CNN-based object detection algorithms, these algorithms can be categorised into three groups.

- Two-Stage detectors
- One-Stage detectors
- Advanced detectors

✓ **Two-Stage detectors**

Two-stage detectors perform object detection by sequentially processing images in two stages, starting with a coarser analysis and then refining it to achieve higher accuracy in the algorithm. Two-stage object detection algorithms employ a multi-step methodology. During the initial phase, they suggest potential regions of objects using a Region Proposal Network (RPN) or a comparable mechanism. The proposed regions undergo classification and refinement in the second stage to achieve the final object detection.

The development of object detection frameworks has seen significant advancements, with essential contributions such as the Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2014), which introduced a two-stage process using selective search for Region Proposals and CNNs for feature extraction and classification. Spatial Pyramid Pooling (SPPNet) (He et al., 2014) improved deep convolutional networks' performance by enabling the processing of random-sized images and producing fixed-length feature vectors. Fast R-CNN (Girshick, 2015)streamlined the process by unifying object classification and bounding box regression in a single-stage network, introducing the Region of Interest (RoI) pooling layer for efficient feature map extraction. Building upon this, Faster R-CNN (Ren et al., 2015) introduced Region Proposal Networks (RPN) to generate high-quality proposals, removing the need for external methods. Mask R-CNN (He et al., 2017)extended Faster R-CNN by incorporating a parallel branch for pixel-level object mask prediction, allowing accurate segmentation.

Deformable Convolutional Networks (DCN) (Dai et al., 2017) introduced adaptive receptive fields, handling geometric variations effectively. Feature Pyramid Network (FPN) (Lin, Dollár, et al., 2017) incorporated squeeze-and-excitation modules for real-time object detection, while Deformable ConvNets v2 (Zhu et al., 2019)enhanced spatial transformation handling. NAS-FPN (Ghiasi et al. Brain, 2019) utilised neural architecture search for optimal feature pyramid generation. Deformable DETR (Zhu et al., 2020) employed deformable transformers for flexible object structure modelling, and UFOD (UFOD – an open-source library) provided a user-friendly interface for automated model selection and hyperparameter optimisation (García-Domínguez et al., 2021). Sparse R-CNN (Sun et al., 2021) innovatively combined proposal generation and detection stages, utilising sparse convolutional layers for adaptive sparse proposals and reduced computational burden.

✓ **One-Stage Detectors** One-stage object detection algorithms perform object detection in a *'single step'* without requiring explicit region proposal. They directly predict object locations and classes on a dense grid of candidate bounding boxes, reducing the computational time in object detection.

You Only Look Once (YOLO) (Redmon et al., 2016), a one-stage detector, revolutionised object detection by treating it as a single regression problem. It accomplished remarkable speed and accuracy by predicting bounding boxes and class probabilities in a single network pass. Single Shot Multibox Detector (SSD) (W. Liu et al., 2015) excels in real-time detection through multi-scale feature maps and anchor boxes, providing accurate results across different object scales and aspect ratios. RetinaNet (Lin, Goyal, et al., 2017) addressed class imbalance using focal loss, focusing on challenging instances and reducing the dominance of well-classified backgrounds. SqueezeDet (B. Wu et al., 2017) introduced a compact and efficient fully convolutional network with squeeze-and-excitation modules for real-time object detection in autonomous driving scenarios.

CenterNet (Duan et al., 2019)innovatively represents objects as single points, enhancing computation speed and localisation accuracy. At the same time, FCOS (Tian et al., 2019) adopts a fully convolutional design, eliminating anchor boxes and introducing a novel focal loss for precise object localisation. EfficientDet (Mingxing et al. Le, 2020) employs compound scaling to optimise architecture, resolution, and depth, covering various resource limitations. YOLOv5 (L. Jiang et al., 2022), designed for traffic sign detection, incorporates a balanced feature pyramid network and attention modules to enhance multi-scale feature capture and improve accuracy. Each framework uniquely contributes to the evolution of object detection methodologies.

✓ **Advanced Detectors** Advanced detectors include state-of-the-art object detection algorithms incorporating improvements over the basic two-stage and one-stage methods. These advancements can involve better backbone architectures, more efficient feature extraction techniques, or new object localisation and recognition strategies.

In object detection, innovative approaches have emerged to redefine the traditional bounding box model. For example, CornerNet (Mingxing et al. Le, 2019) breaks away from conventional bounding boxes by directly predicting object corners as keypoints, employing a unique corner pooling technique and keypoint estimation method for improved accuracy and robustness. Similarly, the Point methodology treats objects as single points (Zhou et al., 2019), removing the complexity of bounding boxes, and utilises keypoint detection to regress object centres, resulting in significantly improved detection accuracy and efficiency. The CenterNet framework, introduced within this context, achieves state-of-the-art performance on various benchmarks while maintaining real-time performance. CornerNet-Lite (Law et al., 2019) improves this concept by focusing on efficiency, representing objects as keypoints, and employing a two-step keypoint regression process to achieve comparable performance while reducing computational complexity.

Another notable advancement is FoveaBox (Kong et al., 2020), which deviates from anchor-based approaches and uses a focal region-based strategy to enhance object localisation and recognition. FoveaBox outperforms anchor-based models using a multi-level feature fusion scheme and the novel concept of ‘foveation’, demonstrating its potential to revolutionise object detection by providing a more effective alternative. Collectively, these novel approaches contribute to the ongoing evolution of object detection methodologies, with the promise of improved accuracy and efficiency in a wide range of practical applications.

Object Detection Algorithms in TIR Videos

Object detection in UAV TIR videos presents unique challenges due to the distinct characteristics of thermal imagery, such as low resolution, complex image background, high noise levels, long imaging distance, variable environmental conditions, flight angles, and lack of publicly labelled datasets and TIR detection for multiple scenes and objects (C. Jiang et al., 2022). Despite the challenges, in recent years, CNN algorithms have shown promising results in addressing these challenges and facilitating accurate detection of objects, including humans and elephants, in TIR videos captured by UAVs.

Various CNN-based algorithms have been developed and adapted specifically for object detection in TIR imagery. These algorithms utilise deep learning architectures to autonomously acquire distinguishing characteristics from unprocessed TIR data, facilitating strong detection capabilities under challenging situations. In this section, we explore some of the prominent CNN-based algorithms tailored for object detection in TIR videos:

✓ **YOLO**

In 2016, Redmon et al. introduced YOLO algorithms, a popular real-time object detection algorithm known for its efficiency and accuracy. The YOLO architecture divides the input image into a grid and predicts bounding boxes and class probabilities directly from the whole image in a single forward pass of the network. This approach enables YOLO to achieve real-time performance, making it well-suited for applications such as UAV-based monitoring of human-elephant conflict in TIR videos.

✓ **SSD**

SSD is another real-time object detection algorithm introduced by W. Liu et al. in 2015 that predicts multiple bounding boxes and class probabilities at different scales and aspect ratios within a single network architecture. By incorporating feature maps from multiple convolutional layers, SSD achieves robustness to object scale variations and maintains high detection accuracy across different object sizes. This makes SSD particularly suitable for detecting elephants and humans in TIR videos captured by UAVs.

✓ **Faster R-CNN**

In 2015, Ren et al. introduced Faster R-CNN is a two-stage object detection framework that first generates region proposals using a Region Proposal Network (RPN) and then refines these proposals through a separate network for classification and bounding box regression. While slightly slower than YOLO and SSD, Faster R-CNN offers superior accuracy and localisation precision (ref table 01). It is well-suited for scenarios where precision is paramount, such as identifying small objects or distinguishing between similar classes.

Table 01: Comparison of Precision over Speed

Algorithm		Data Set	Mean Average Precision (mAP)	Speed (Frames per Second)
YOLO		VOC-2012	63.4%	45 FPS
SSD	SSD300		74.3%	59 FPS
	SSD500		76.9%	22 FPS
Fast RCNN’s			73.2%	7 FPS

Source: Murthy et al., 2020

These CNN-based algorithms have shown significant advancements in object detection accuracy and efficiency in TIR videos, laying the foundation for improved monitoring and management of human-elephant conflict using UAV technology. However, each algorithm has its strengths and weaknesses, and the choice of algorithm depends on factors such as computational resources, real-time processing requirements, and the specific characteristics of the HEC management scenario.

Integration Of CNN Algorithms in UAV Systems

Integrating CNN algorithms for object detection into UAV systems is a critical step towards real-time monitoring and management of HEC. However, this integration poses several challenges related to computational efficiency, onboard processing capabilities, and TIR video streaming optimisation. In this section, we discuss the considerations and strategies for effectively integrating CNN algorithms into UAV systems for HEC management:

✓ **Computational Efficiency**: CNN-based object detection algorithms can be computationally intensive, requiring substantial processing power to perform real-time inference on streaming TIR videos. To address this challenge, researchers have explored techniques such as model quantisation (Z. Liu et al., n.d.), network pruning (Habib et al., n.d.), and hardware acceleration (Du et al., 2024) using specialised processing units (e.g., GPUs or FPGAs) to optimise the computational efficiency of CNN algorithms on UAV platforms. Additionally, designing lightweight CNN architectures optimised for resource-constrained environments can improve the computational efficiency of object detection in UAV systems.

✓ **Onboard Processing Capabilities**: UAVs often have limited onboard processing capabilities, so developing efficient algorithms and architectures for real-time object detection is necessary. By leveraging distributed processing techniques and parallel computing frameworks, such as CUDA or OpenCL, UAV systems can harness the computational power of onboard hardware components to accelerate CNN inference tasks (Jordà et al., 2021). Moreover, deploying hybrid approaches that combine onboard and offboard processing can improve resource utilisation while maintaining low-latency performance in HEC management scenarios.

✓ **Optimization for TIR Video Streaming**: TIR videos captured by UAVs have distinct characteristics, such as low resolution, high noise levels, and variable environmental conditions, which make challenges for object detection algorithms. Researchers have explored data augmentation techniques, domain adaptation strategies, and transfer learning approaches to optimise CNN algorithms for TIR video streaming to improve model robustness and generalisation capabilities in diverse environmental conditions (Zoph et al., 2019). Additionally, integrating sensor fusion techniques, such as combining TIR imagery with visual or LiDAR data, can improve the accuracy and reliability of object detection in UAV systems for HEC management.

By addressing these challenges and considerations, integrating CNN algorithms into UAV systems can significantly enhance the effectiveness of HEC mitigation efforts. Real-time monitoring and early detection of human and elephant presence enable proactive intervention strategies, such as alerting local authorities or deploying deterrents, to prevent conflicts and minimise risks to both human and elephant populations.

Evaluation Metrics and Performance Analysis

Assessing object detection algorithms' performance in UAV-based HEC monitoring requires appropriate evaluation metrics and comprehensive performance analysis. In this section, we describe the metrics commonly used to evaluate object detection algorithms, conduct a comparative study of three different CNN algorithms, namely YOLO, SSD, and Faster R-CNN, and discuss the trade-offs between accuracy and processing speed in the context of emergency response requirements. Evaluating object detection algorithms involves several key metrics that help determine their performance and effectiveness. Here are four commonly used metrics:

✓ **Evaluation Metrics**:

- *Precision*: The ratio of correctly predicted positive observations to the total predicted positives. It indicates how many of the detected objects are actually relevant.

- *Recall*: The ratio of correctly predicted positive observations to all observations in the actual class. It measures the algorithm's ability to detect all relevant objects.
- *F1 Score*: The harmonic mean of precision and recall. It provides a single metric that balances both precision and recall.
- *Intersection over Union (IoU)*: Measures the overlap between the predicted and ground truth bounding boxes. Higher IoU indicates better localisation accuracy.

Table 02: A comparison of the evaluation metrics with CNN algorithms

CNN Algorithm	**Precision**	**Recall**	**F1 Score**	**IoU**
YOLO	High precision	Lower compared to other algorithms	Higher than YOLO	Performs well
SSD	Good balance between precision and recall	SSD has better recall compared to YOLO	Generally higher than YOLO	Achieves good IoU
Faster R-CNN	High precision	High Recall	The highest among these three algorithms	Excels in IoU

✓ **Performance Analysis**:

- *Comparative Analysis*: A comparative analysis of different CNN algorithms, such as YOLO, SSD, and Faster R-CNN, evaluates their performance across various metrics on HEC-related datasets and real-world scenarios (ref Table 01). This analysis reveals the strengths and weaknesses of each algorithm in detecting humans and elephants in thermal infrared (TIR) videos captured by UAVs.
- *Real-world Scenario Evaluation*: Evaluating the performance of object detection algorithms in real-world HEC management scenarios involves deploying integrated UAV systems equipped with CNN algorithms and evaluating their effectiveness in detecting and tracking relevant objects in streaming TIR videos. This evaluation considers detection accuracy, false

positive rates, and processing speed under different environmental and operational constraints.

Table 03: A comparative study of CNN algorithms used for object detection

CNN Algorithm	Strengths	Weaknesses	Use Cases
YOLO	Real-time detection, high speed, and efficiency	Lower accuracy compared to some other models, especially for small objects	Applications requiring real-time detection, such as autonomous driving and surveillance
SSD	A balance between speed and accuracy, efficient multi-scale feature maps	Slightly lower accuracy than Faster R-CNN	Applications needing a good trade-off between speed and accuracy, like mobile and embedded systems
Faster R-CNN	High accuracy and robust performance for various object sizes	Slower inference time compared to YOLO	Scenarios where accuracy is more critical than speed, such as medical imaging and detailed image analysis

✓ **Trade-offs between Accuracy and Processing Speed**: Balancing accuracy and processing speed is crucial in emergency response situations, where early detection and intervention are critical. While Faster R-CNN may offer higher accuracy, they often require more computational resources and have longer inference times than faster algorithms like YOLO and SSD. Therefore, selecting the most suitable algorithm depends on the specific requirements of the HEC management scenario, considering factors such as response time, resource availability, and detection performance trade-offs.

Researchers and practitioners can gain valuable insights into the effectiveness of object detection algorithms in UAV-based HEC monitoring by using appropriate evaluation metrics and conducting a thorough performance analysis. This analysis informs decision-making processes and makes it easier to select and optimise algorithms for real-world deployment, ultimately contributing to successfully mitigating human-elephant conflict and promoting coexistence between humans and elephants in conflict-prone regions.

Future Directions and Challenges

Several future directions and challenges emerge as the object detection field in UAV-based monitoring of HEC continues to evolve, offering new opportunities for innovation and advancement. In this section, we identify emerging trends, potential advancements, and critical challenges in object detection technology for HEC mitigation:

✓ **Trends**

- *Multi-Sensor Fusion*: Integrating data from multiple sensors, such as TIR cameras, visual cameras, and LiDAR sensors, can potentially improve the accuracy and reliability of object detection algorithms in diverse environmental conditions. Researchers can overcome limitations inherent in individual sensors and improve the robustness of detection systems in challenging scenarios, such as dense foliage or adverse weather conditions, by combining information from various sensor systems.
- *Semi-supervised Learning:* Semi-supervised learning techniques can help train object detection algorithms with limited annotated data, often a bottleneck in HEC management applications. Semi-supervised learning algorithms can effectively bootstrap the training process and improve detection performance by combining a small amount of labelled data with a large pool of unlabelled data, especially when collecting large, labelled datasets, which is impractical or prohibitively expensive.
- *Adaptive Algorithms for Dynamic Environments*: Developing adaptive object detection algorithms that dynamically adjust their parameters and strategies in response to changing environmental conditions and operational requirements is essential for real-world deployment in dynamic HEC scenarios. Adaptive algorithms can automatically optimise their performance in response to changes in terrain, vegetation density, and animal behaviour, increasing their effectiveness and adaptability in monitoring and managing human-elephant conflict.

✓ **Challenges**

- *Ethical and Practical Considerations*: Deploying AI-based UAVs for wildlife management and emergency response raises ethical and practical concerns about privacy, data security, and community engagement.

Transparency, accountability, and stakeholder participation in developing and deploying UAV-based solutions are critical for establishing trust and addressing potential concerns among local communities and conservation stakeholders.

- *Scalability and Generalization*: Scalability and generalisation of object detection algorithms across diverse geographical regions, elephant populations, and environmental conditions remain significant challenges. Developing algorithms that can generalise effectively to new environments and elephant behaviours while maintaining high detection accuracy is essential for the widespread adoption and long-term viability of UAV-based HEC mitigation strategies.
- *Regulatory and Policy Frameworks*: Navigating regulatory and policy frameworks that govern the use of UAVs in wildlife management and conservation presents logistical and legal challenges. Establishing clear guidelines, protocols, and regulations for the ethical and responsible deployment of UAVs in HEC mitigation efforts is essential for ensuring compliance with local regulations, wildlife protection, and protecting human rights.

Addressing these challenges and capitalising on emerging trends in object detection technology will pave the way for innovative solutions to human-elephant conflict and promote long-term coexistence between humans and elephants in conflict-prone areas.

Conclusion

To summarise, integrating CNN algorithms into UAV video streaming represents a transformative approach in addressing the complex challenges of HEC mitigation. Researchers and practitioners have made significant progress in improving the effectiveness of real-time monitoring and management strategies in conflict-prone areas by leveraging advanced technologies and innovative methodologies.

This review has explored the evolution of CNN algorithms from traditional to deep-learning object detection methods and further elaborated on the two-stage, one-stage, and advanced deep-learning methods used in the present-day context. Moreover, the role of CNN-based algorithms, such as YOLO, SSD, and Faster R-CNN, has been discussed, enabling accurate and efficient detection of humans and elephants in challenging environmental conditions.

Furthermore, the considerations and challenges associated with integrating CNN algorithms into UAV systems, such as computational efficiency, onboard processing capabilities, and optimisation for TIR video streaming, have been investigated. Also, Researchers can improve the scalability, adaptability, and effectiveness of UAV-based HEC mitigation strategies by addressing these challenges and utilising emerging technologies.

It is imperative to continue developing object detection technology for HEC management while considering ethical, regulatory, and policy considerations. By promoting interdisciplinary collaborations, encouraging stakeholder engagement, and embracing responsible deployment practices, it can develop holistic solutions prioritising wildlife conservation, human safety, and community well-being.

In conclusion, this review emphasises the importance of CNN-based object detection in UAV TIR video streaming for addressing human-elephant conflict. By leveraging technology and innovation, we can pave the way for long-term coexistence between humans and elephants, ensuring a brighter future for both species and the ecosystems they inhabit.

Reference

Akula, A., Khanna, N., Ghosh, R., Kumar, S., Das, A., & Sardana, H. K. (2014). Adaptive contour-based statistical background subtraction method for moving target detection in infrared video sequences. *Infrared Physics and Technology*, *63*, 103–109. https://doi.org/10.1016/j.infrared.2013.12.012

Anuradha, J. M. P. N., Fujimura, M., Inaoka, T., & Sakai, N. (2019). The role of agricultural land use pattern dynamics on elephant habitat depletion and human-elephant conflict in Sri Lanka. *Sustainability (Switzerland)*, *11*(10). https://doi.org/10.3390/su11102818

Billah, M. M., Rahman, M. M., Abedin, J., & Akter, H. (2021). Land cover change and its impact on human–elephant conflict: a case from Fashiakhali forest reserve in Bangladesh. *SN Applied Sciences*, *3*(6). https://doi.org/10.1007/s42452-021-04625-1

Castaldo-Walsh, C. (2019). *NSUWorks Human-Wildlife Conflict and Coexistence in a More-than-Human World: A Multiple Case Study Exploring the Human-Elephant-Conservation Nexus in Namibia and Sri Lanka*. https://nsuworks.nova.edu/shss_dcar_etd

D'Acremont, A., Fablet, R., Baussard, A., & Quin, G. (2019). CNN-based target recognition and identification for infrared imaging in defense systems. *Sensors (Switzerland)*, *19*(9). https://doi.org/10.3390/s19092040

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). *Deformable Convolutional Networks*. https://github.com/

Dalal, N., Histograms, B. T., & Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. 886–893. https://doi.org/10.1109/CVPR.2005.177ï

Das, B. J., Saikia, B. N., Baruah, K. K., Bora, A., & Bora, M. (2014). Nutritional evaluation of fodder, its preference and crop raiding by wild Asian elephant (Elephas maximus) in Sonitpur district of Assam, India. *Veterinary World*, *7*(12), 1082–1089. https://doi.org/10.14202/vetworld.2014.1082-1089

Du, D., Gong, G., & Chu, X. (2024). *Model Quantization and Hardware Acceleration for Vision Transformers: A Comprehensive Survey*. http://arxiv.org/abs/2405.00314

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). *CenterNet: Keypoint Triplets for Object Detection*. https://github.com/

Felzenszwalb, P. F., Girshick, R. B., & Mcallester, D. (2008). *Cascade Object Detection with Deformable Part Models* *.

Fernando, P. (2015). Managing elephants in Sri Lanka: where we are and where we need to be. *Ceylon Journal of Science (Biological Sciences)*, *44*(1), 1–11. https://doi.org/10.4038/cjsbs.v44i1.7336

Fernando, P., De Silva, M. K. C. R., Jayasinghe, L. K. A., Janaka, H. K., & Pastorini, J. (2021). First country-wide survey of the Endangered Asian elephant: Towards better conservation and management in Sri Lanka. *ORYX*, *55*(1), 46–55. https://doi.org/10.1017/S0030605318001254

García-Domínguez, M., Domínguez, C., Heras, J., Mata, E., & Pascual, V. (2021). UFOD: An AutoML framework for the construction, comparison, and combination of object detection models. *Pattern Recognition Letters*, *145*, 135–140. https://doi.org/10.1016/j.patrec.2021.01.022

George Wittemyer, P. E. W. T. B. A. C. O. B. J. S. B. (2008). Accelerated Human Population Growth at Protected Area Edges. *Science*, *321*(5885), 123–126. https://doi.org/10.1126/science.1154449

Ghiasi, G., Lin, T.-Y., & Le Google Brain, Q. V. (2019). *NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection*.

Girshick, R. (2015). *Fast R-CNN*. https://github.com/rbgirshick/

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. http://arxiv.

Gonzalez, L. F., Montes, G. A., Puig, E., Johnson, S., Mengersen, K., & Gaston, K. J. (2016). Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors (Switzerland)*, *16*(1). https://doi.org/10.3390/s16010097

Habib, G., Saleem, J., & Lall, B. (n.d.). *Knowledge Distillation in Vision Transformers: A Critical Review*.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*.

He, K., Zhang, X., Ren, S., & Sun, J. (2014). *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*. https://doi.org/10.1007/978-3-319-10578-9_23

Hoare, R. E. (1999). Determinants of human-elephant conflict in a land-use mosaic. *Journal of Applied Ecology*, *36*(5), 689–700. https://doi.org/10.1046/j.1365-2664.1999.00437.x

Jiang, C., Ren, H., Ye, X., Zhu, J., Zeng, H., Nan, Y., Sun, M., Ren, X., & Huo, H. (2022). Object detection from UAV thermal infrared images and videos using YOLO models. *International Journal of Applied Earth Observation and Geoinformation*, *112*. https://doi.org/10.1016/j.jag.2022.102912

Jiang, L., Liu, H., Zhu, H., & Zhang, G. (2022). Improved YOLO v5 with balanced feature pyramid and attention module for traffic sign detection. *MATEC Web of Conferences*, *355*, 03023. https://doi.org/10.1051/matecconf/202235503023

Jordà, M., Valero-Lara, P., & Peña, A. J. (2021). *cuConv: A CUDA Implementation of Convolution for CNN Inference*. https://doi.org/10.1007/s10586-021-03494-y

Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L., & Shi, J. (2020). FoveaBox: Beyound Anchor-Based Object Detection. *IEEE Transactions on Image Processing*, *29*, 7389–7398. https://doi.org/10.1109/TIP.2020.3002345

Köpke, S., Withanachchi, S. S., Pathiranage, R., Withanachchi, C. R., Udayakanthi, T. G. D., Nissanka, N. M. T. S., Warapitiya, C., Nissanka, L. N. A. B. M., Ranasinghe, R. A. N. N., Senarathna, T. M. C. D., Schleyer, C., & Thiel, A. (2021). Human—elephant conflict in Sri Lanka: A critical review of causal explanations. *Sustainability (Switzerland)*, *13*(15). https://doi.org/10.3390/su13158625

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Law, H., Teng, Y., Russakovsky, O., & Deng, J. (2019). *CornerNet-Lite: Efficient Keypoint Based Object Detection*. http://arxiv.org/abs/1904.08900

Leimgruber, P., Oo, Z. M., Aung, M., Kelly, D. S., Wemmer, C., Senior, B., & Songer, M. (2011). Current Status of Asian Elephants in Myanmar. In *Gajah* (Vol. 35).

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature Pyramid Networks for Object Detection*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal Loss for Dense Object Detection*.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). *SSD: Single Shot MultiBox Detector*. https://doi.org/10.1007/978-3-319-46448-0_2

Liu, Z., Wang, Y., Han, K., Zhang, W., Ma, S., & Gao, W. (n.d.). *Post-Training Quantization for Vision Transformer*. https://gitee.com/mindspore/models/tree/master/research/cv/VT-

Mingxing Tan, Ruoming Pang, & Quoc V. Le. (2020). *EfficientDet: Scalable and Efficient Object Detection*. https://github.com/google/

Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. In *Image and Vision Computing* (Vol. 104). Elsevier Ltd. https://doi.org/10.1016/j.imavis.2020.104046

Murthy, C. B., Hashmi, M. F., Bokde, N. D., & Geem, Z. W. (2020). Investigations of object detection in images/videos using various deep learning techniques and embedded platforms-A comprehensive review. In *Applied Sciences (Switzerland)* (Vol. 10, Issue 9). MDPI AG. https://doi.org/10.3390/app10093280

Nasrabadi, N. M. (2019). DeepTarget: An Automatic Target Recognition Using Deep Convolutional Neural Networks. *IEEE Transactions on Aerospace and Electronic Systems*, *55*(6), 2687–2697. https://doi.org/10.1109/TAES.2019.2894050

Naughton-Treves, L. (2010). Farming the forest edge: vulnerable places and people around Kibale national park, Uganda. *Geographical Review*, *87*(1), 27–46. https://doi.org/10.1111/j.1931-0846.1997.tb00058.x

Nyhus, P. J. (2016). Human-Wildlife Conflict and Coexistence. *Annual Review of Environment and Resources*, *41*, 143–171. https://doi.org/10.1146/annurev-environ-110615-085634

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. http://pjreddie.com/yolo/

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. https://github.com/

Santiapillai, C., Wijeyamohan, S., Bandara, G., Athurupana, R., Dissanayake, N., & Read, B. (2010). An assessment of the human-elephant conflict in Sri Lanka. In *J. Sci. (Bio. Sci.)* (Vol. 39, Issue 1).

Shaffer, L. J., Khadka, K. K., Van Den Hoek, J., & Naithani, K. J. (2019). Human-elephant conflict: A review of current management strategies and future directions. In *Frontiers in Ecology and Evolution* (Vol. 6, Issue JAN). Frontiers Media S.A. https://doi.org/10.3389/fevo.2018.00235

Sitati, N. W., Walpole, M. J., Smith, R. J., & Leader-Williams, N. (2003). Predicting spatial aspects of human-elephant conflict. *Journal of Applied Ecology*, *40*(4), 667–677. https://doi.org/10.1046/j.1365-2664.2003.00828.x

Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C., & Luo, P. (2021). *Sparse R-CNN: End-to-End Object Detection with Learnable Proposals*. https://github.com/PeizeSun/SparseR-CNN

Tian, Z., Shen, C., Chen, H., & He, T. (2019). *FCOS: Fully Convolutional One-Stage Object Detection*.

Tiller, L. N., Humle, T., Amin, R., Deere, N. J., Lago, B. O., Leader-Williams, N., Sinoni, F. K., Sitati, N., Walpole, M., & Smith, R. J. (2021). Changing seasonal, temporal and spatial crop-raiding trends over 15 years in a human-elephant conflict hotspot. *Biological Conservation*, *254*. https://doi.org/10.1016/j.biocon.2020.108941

Viola, P., & Jones, M. (2004). *Merl-a Mitsubishi electric research laboratory Rapid Object Detection Using a Boosted Cascade of Simple Features Rapid Object Detection using a Boosted Cascade of Simple Features*. http://www.merl.com

Witczuk, J., Pagacz, S., Zmarz, A., & Cypel, M. (2018). Exploring the feasibility of unmanned aerial vehicles and thermal imaging for ungulate surveys in forests - preliminary results. *International Journal of Remote Sensing*, *39*(15–16), 5504–5521. https://doi.org/10.1080/01431161.2017.1390621

Wu, B., Iandola, F., Jin, P. H., & Keutzer, K. (2017). *SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving*. https://blogs.nvidia.com/blog/2016/09/28/

Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*, 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

Zhou, X., Wang, D., & Krähenbühl, P. (2019). *Objects as Points*. http://arxiv.org/abs/1904.07850

Zhu, X., Hu, H., Lin, S., & Dai, J. (2019). *Deformable ConvNets v2: More Deformable, Better Results*.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2020). *Deformable DETR: Deformable Transformers for End-to-End Object Detection*. http://arxiv.org/abs/2010.04159

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., & Le, Q. V. (2019). *Learning Data Augmentation Strategies for Object Detection*. http://arxiv.org/abs/1906.11172