A Comprehensive Review of Deep Learning Approaches for Animal Detection on Video Data

Prashanth Kumar, Suhuai Luo, Kamran Shaukat
School of Information and Physical Sciences
The University of Newcastle
Newcastle, Australia

Abstract—**Integrating deep learning techniques into computer vision application has ushered in a new era of automated analysis and interpretation of visual data. In recent years, a surge of interest has been witnessed in applying these methodologies towards detecting animals in video streams, promising transformative impacts on diverse fields such as ecology and agriculture. This paper presents an extensive and meticulous review of the latest deep-learning approaches employed for animal detection in video data. This study looks closely at ways to detect animals in videos using deep learning. This study explores various Deep learning methods for detecting many animals in multiple environments. The analysis also pays close attention to preparing the data, picking out important features, and reusing what has been learned from one task to help with another. In addition to highlighting successful methodologies, this review addresses the challenges and limitations inherent in these approaches issues such as limited data availability and adapting to technological advancements present significant hurdles. Recognising and understanding these challenges is crucial in shaping the future focus of research endeavours. Thus, this comprehensive review is an indispensable tool for anyone keen on employing these potent computer methods for animal detection in videos. It takes the latest ideas and shows where study can explore further to improve them. Furthermore, this comprehensive review has demonstrated that a more sustainable and balanced relationship between humans and animals can be achieved by harnessing the power of deep learning in animal detection. This research contributes to computer vision and holds immense promise in safeguarding biodiversity and promoting responsible land use practices, especially within agricultural domains. The insights from this study propel us towards a future where advanced technology and ecological harmony go hand in hand, ultimately benefiting both humans and the animal kingdom. The survey aims to provide a comprehensive overview of the cutting-edge developments in applying deep learning models for animal detection through cameras by elucidating the significance of these techniques in advancing the accuracy and efficiency of animal detection processes.**

Keywords—Machine learning; deep learning; animal detection; convolutional neural networks; video-based; deep learning models

I. INTRODUCTION

Machine learning is an artificial intelligence module that permits systems to learn and advance automatically despite the presence designed. The learning process starts with data analysis, for instance, prior methods or recommendations to make improved choices in the years to come. The foremost goal is to permit programs to teach themselves without human involvement or help and to correct their errors through this learning. Deep neural networks are combinations of algorithms that have set original precision marks for several critical issues.

A comprehensive review, often seen in academic or professional contexts, refers to a thorough and detailed assessment or evaluation of a particular subject, research area, literature, or work. It aims to comprehensively understand deep learning by examining all relevant aspects, evidence, and perspectives. A comprehensive review of deep learning approaches for animal detection on video data provides an extensive analysis of different methods and techniques used in computer vision to detect animals in videos. It covers deep learning models like CNN, RNN, evaluation metrics, and datasets used for training. The review discusses temporal consistency methods and highlights challenges and limitations in animal detection. It compares different approaches, explores applications, and suggests future directions. Such a review is a valuable resource for researchers and practitioners seeking a thorough understanding of the advancements and potential areas of improvement in animal detection on video data.

Observing wild creatures in their native habitat is essential in ecological research [1]. Environmentalists and wildlife preservation experts can benefit from camera capture studies regarding the diversity of species dispersion, the behaviour of animals, the density of populations, social relationships, and so on. Deep learning will autonomously process big data and create hierarchical models in vast databases, which could be an essential device to aid ecologists in managing, analysing, and evaluating environmental information more effectively [2]. Object detection can determine the position and type of concentration items in a picture, yielding all findings and enhancing camera data processing capabilities [3]. Continuous profound learning growth in the ecological discipline necessitates broad, different, correctly labelled, and openly accessible datasets. In some datasets, the makeup of various species could be more balanced [4]. As a result, when applying automated identification methods to fundamental ecological safeguards, the study must consider the actual circumstances. Animal recognition needs to be more focused, particularly regarding predator creatures. Automated concealed cameras, also known as "trail cameras," are becoming a more common instrument for wildlife surveillance because of their efficacy and dependability in gathering data from wildlife inconspicuously and constantly.

However, outdoors, deploying a trail camera device presents several obstacles, such as dealing with low light conditions, small images, or network limitations. Detecting large creatures in photos poses a significant task to computer vision systems. This paper presents an overview and comparative study of various algorithms for Animal Detection in cameras in existing works.

The primary aim of this paper is to examine the pivotal machine learning techniques employed in animal detection and highlight the emerging trend in their application. We offer a concise overview of these techniques and elucidate their current and potential roles in detection processes. Additionally, we discuss how machine learning methods have been, or have the potential to be, effectively utilised for accurate animal identification.

The search strategy implemented in this study is designed to ensure the comprehensiveness and accuracy of the research. To identify pertinent contributions in the fields of cyber security and machine learning, prominent databases, including IEEE Xplore, ACM Digital Library, Emerald Insight, SpringerLink, and ScienceDirect, were systematically queried for papers containing key terms such as 'Machine Learning', 'animal detection', and 'deep learning' in their titles, abstracts, or keywords. Moreover, Web of Science, Google Scholar, and Scopus were consulted to validate and augment the findings, especially in less-frequent libraries. Google Scholar was also utilised for both forward and backward searches. Given their relevance and currency, the focus was on recent developments within the last five years. These online repositories were chosen due to their extensive coverage of peer-reviewed full-text journals, conference proceedings, book chapters, and machine learning and cyber security reports. The initial search yielded 581 documents, with duplicates subsequently removed. Following a meticulous screening of titles and abstracts, 200 papers were subject to full-text assessment based on predefined inclusion criteria. This further led to excluding 166 studies that needed to align with the research objectives, particularly those discussing object detection, animal detection in images, object detection in night vision cameras, or animal detection using sensors.

Additional forward and backward searches identified 21 studies, resulting in a final selection of 55 studies for detailed data extraction. Fig. 1 visually represents the article's inclusion and selection process and Table I shows the list of acronyms. Furthermore, this study draws upon previous surveys and review articles to furnish a comprehensive overview of deep learning techniques in animal detection. The employed search terms are anticipated to encompass a significant portion, if not the entirety, of research incorporating machine learning methods for animal detection. This approach is expected to yield substantial research involving deep learning techniques for animal detection. Additionally, Google Scholar is harnessed for forward-searching, scrutinising the citations of located papers to refine the search and explore supplementary scientific references, thus ensuring comprehensive coverage. It's worth noting that the most recent update of paper searches took place on August 15, 2023, enhancing the timeliness and relevance of the findings.

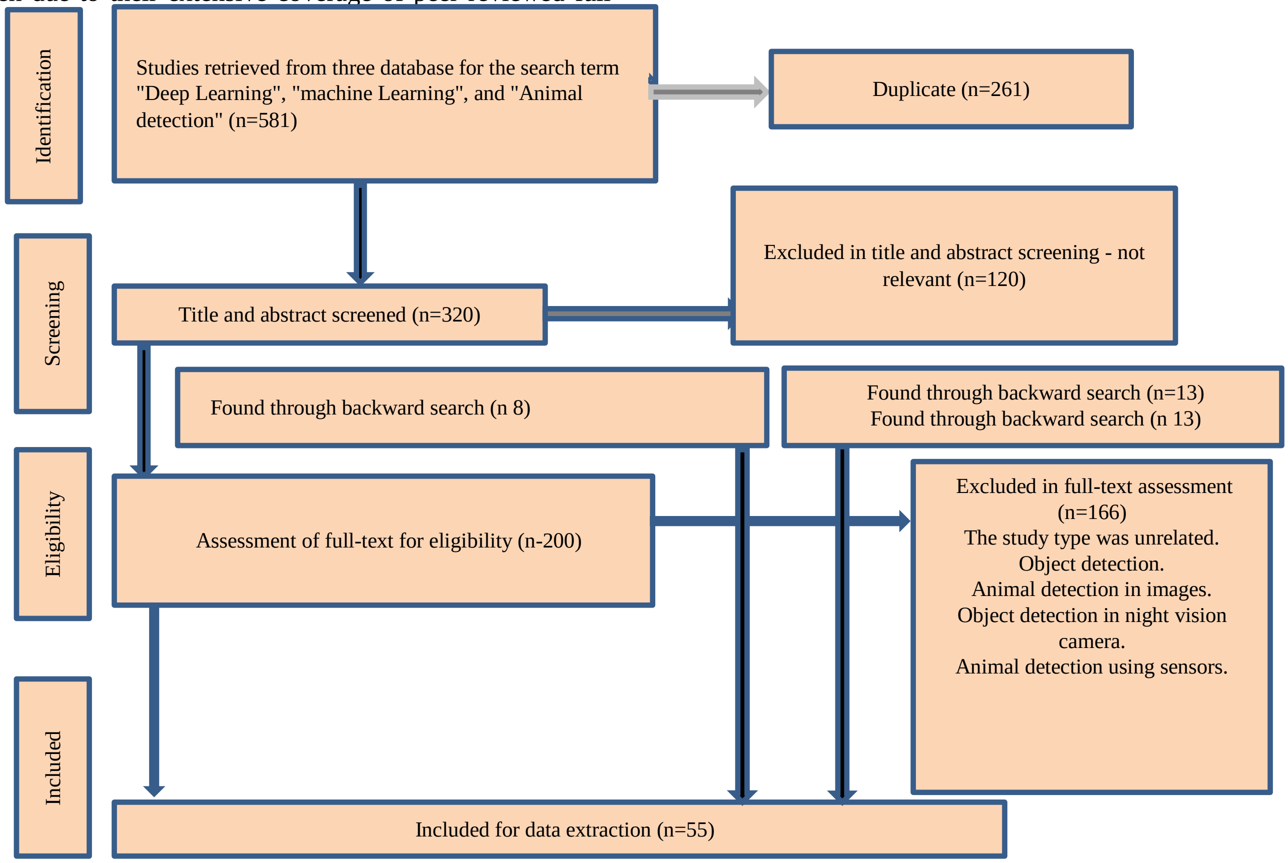
Identification
Studies retrieved from three database for the search term "Deep Learning", "machine Learning", and "Animal detection" (n=581)
Duplicate (n=261)
Screening
Title and abstract screened (n=320)
Excluded in title and abstract screening - not relevant (n=120)
Found through backward search (n 8)
Found through backward search (n=13)
Found through backward search (n 13)
Eligibility
Assessment of full-text for eligibility (n-200)
Excluded in full-text assessment (n=166)
The study type was unrelated.
Object detection.
Animal detection in images.
Object detection in night vision camera.
Animal detection using sensors.
Included
Included for data extraction (n=55)

Fig. 1. An illustrative view of the process for article selection.

TABLE I. LIST OF ACRONYMS

FNNs	Feedforward Neural Networks
CNNs	Convolutional Neural Networks
RNNs	Recurrent Neural Networks
LSTMs	Extended Short-Term Memory Networks
GRUs	Gated Recurrent Units
VAEs	Variational Autoencoders
GANs	Generative Adversarial Networks
DBNs	Deep Belief Networks
NLP	Natural Language Processing
ML	Machine learning
DL	Deep Learning
BERT	Bidirectional Encoder Representations from Transformers
CapsNet	Capsule Networks
NEAT	Neuroevolution of Augmenting Topologies
SSD	Single Shot MultiBox Detector
TCNN	Temporal CNN
YOLO	You Only Look Once
R-CNN	Regional Convolutional Neural Networks
RPN	Region proposal networks
STSN	Spatio-Temporal Snippet Network
STAM	Spatio-Temporal Attention Mechanism
Trajnet	Trajectory forecasting
AI	Artificial Intelligence
HOG	Histogram of Oriented Gradients
SVM	Support Vector Machine

Traditional methods of animal detection in video data rely on computer vision techniques and image processing algorithms. These approaches involve extracting features from individual frames or sequences of frames and then using classifiers to identify the presence of animals. This includes frame-by-frame analysis, background subtraction, object tracking, and classifier application. However, these methods face several challenges. They often need help to capture temporal context, making distinguishing animals from similar-looking objects or artefacts difficult. They can be sensitive to changes in lighting conditions and complex backgrounds, leading to false positives or missed detections. Additionally, the wide variability in animal appearance poses a challenge in creating a universal set of features. Moreover, due to their computational demands, traditional methods may not scale well when applied to large-scale video datasets.

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionised animal detection in video data. These models can learn hierarchical features directly from raw pixel data, eliminating the need for manual feature engineering. This enables the network to adapt to a wide range of animal appearances. Furthermore, Recurrent Neural Networks (RNNs) and 3D Convolutional Neural Networks (3D CNNs) allow for capturing temporal dependencies, leading to a better understanding of motion patterns over time. Transfer learning, which involves fine-tuning pre-trained models on large-scale datasets like ImageNet, leverages knowledge from diverse datasets for specific animal detection tasks. Deep learning models are also adept at distinguishing animals from complex backgrounds by automatically extracting relevant features. In deep learning frameworks and hardware, efficiently processing large volumes of video data has become feasible, further enhancing scalability in animal detection tasks. Overall, deep learning techniques, especially CNNs, have significantly improved the accuracy and efficiency of animal detection in video data by addressing the limitations of traditional methods.

II. DEEP LEARNING ARCHITECTURES FOR ANIMAL DETECTION ON VIDEO DATA

Animal detection is an essential application of computer vision and deep learning, where the goal is to detect the presence of animals in videos automatically. Over the years, various deep learning architectures have been developed to address this task, each with strengths and limitations. Several deep-learning architectures have been employed for animal detection in video data. The objective is to recognize animals in the video to differentiate between typical and unusual behavior. Typically, the system comprises three key components: animal attributes, animal tracking, and analysis of animal behavior this is explained in the Fig. 2.

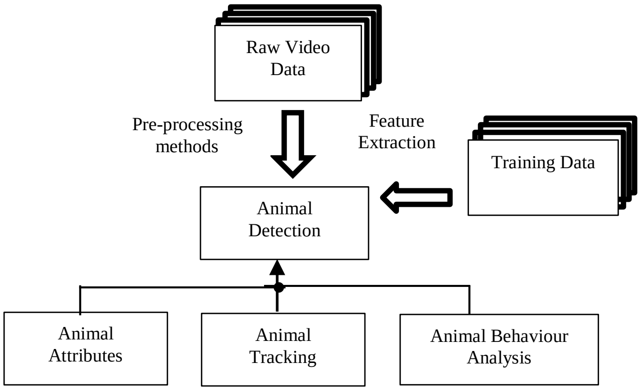

Fig. 2. Flow and structure of animal anomaly detection.

Feedforward Neural Networks (FNNs) represent a foundational paradigm within artificial neural networks. Its hallmark characteristics lies in the unidirectional flow of information, initiating at the input nodes, traversing through a series of hidden layers, and culminating at the output nodes [54]. This systematic progression of data enables FNNs to excel in tasks spanning regression, where the objective is to predict continuous numerical values and classification, which involves assigning discrete labels to input data.

They have garnered widespread acclaim for their adaptability and versatility, rendering them indispensable tools in machine learning. Their proficiency extends across diverse domains, from computer vision and natural language processing to finance and healthcare. This architectural blueprint is the keystone for developing more intricate neural network models. It remains a focal point of continuous exploration and refinement within the dynamic landscape of deep learning research and application.

Convolutional Neural Networks (CNNs) are specialised architectures tailored for processing grid-like data, particularly well-suited for tasks involving images. These networks are distinguished by utilising convolutional layers, enabling them to acquire hierarchical features from the input data autonomously [43]. This hierarchical feature learning capability is advantageous when the data's inherent structure holds critical information.

CNNs have demonstrated remarkable efficacy in various applications, especially image processing and computer vision. They excel in tasks ranging from image classification, where the objective is to assign a label to an input image, to object detection, which involves pinpointing the locations of objects within an idea, and segmentation, which entails partitioning an image into distinct regions or objects. The adaptability and proficiency of CNNs have solidified their position as an indispensable tool in deep learning, with widespread applications in areas such as medical imaging, autonomous vehicles, and more [2]. Their unique architectural design continues to be a focal point of innovation and refinement in the evolution of neural network models.

Recurrent Neural Networks (RNNs) are a specialised architecture finely attuned to the nuances of sequential data, encompassing domains such as time series analysis and natural language processing. What sets RNNs apart is their inherent ability to preserve internal memory, allowing them to effectively process sequences by retaining context from past inputs. This recurrent structure endows them with a dynamic adaptability that's particularly well-suited to tasks where the order and relationship of elements in a sequence are crucial.

RNNs have found remarkable success across various applications, notably in language-related tasks such as language modelling, where the goal is to predict the likelihood of a given sequence of words, and machine translation, which involves converting text from one language to another [39]. Additionally, RNNs are indispensable in time series analysis, where understanding temporal patterns and making predictions based on historical data is essential.

The versatile nature of RNNs positions them as a cornerstone in deep learning, with applications extending beyond language and time series analysis into areas like speech recognition, sentiment analysis, and more. Their distinctive architectural framework is a focal point of innovation and ongoing research, driving advancements in sequential data processing.

Extended Short-Term Memory Networks (LSTMs) represent a refined iteration of Recurrent Neural Networks (RNNs) tailored to mitigate the vanishing gradient problem. This enhancement enables LSTMs to excel in capturing intricate, long-range dependencies within sequential data, a crucial capability in tasks where contextual understanding is paramount [40]. Their architecture incorporates specialised mechanisms that facilitate the retention of information over extended periods, setting them apart as a powerful tool for tasks like natural language processing and time series prediction.

Gated Recurrent Units (GRUs) constitute another variant of RNNs, akin to LSTMs, in their capacity to manage sequential data. Their computational efficiency sets GRUs apart, offering a more streamlined approach to processing sequences while maintaining a similar level of effectiveness. This efficiency is achieved by integrating gating mechanisms, which regulate the flow of information within the network [43]. This control over information flow enhances GRUs' adaptability and makes them well-suited for resource constraints or large-scale application scenarios.

Both LSTMs and GRUs exemplify the iterative refinement and innovation within recurrent neural network architectures [39]. Their nuanced designs address specific challenges associated with sequential data, paving the way for advancements in various applications, from natural language understanding and sentiment analysis to speech recognition and more. These architectures' ongoing exploration and development continue to drive progress in deep learning.

Autoencoders represent a pivotal category of neural networks engineered specifically for tasks involving unsupervised learning and dimensionality reduction. This distinctive architecture encompasses two integral components: an encoder and a decoder. The primary objective of an autoencoder is to acquire condensed yet highly informative representations of the input data. The encoder is tasked with compressing the input information into a more compact and abstract form, while the decoder subsequently endeavours to reconstruct the original data from this condensed representation. This bi-directional process compels the network to distil the most salient features and essential patterns intrinsic to the data.

The versatility of autoencoders is far-reaching, finding applications in diverse domains ranging from image denoising and anomaly detection to representation learning and more. Their effectiveness in unsupervised settings, where labelled training data may be scarce or unavailable, renders them invaluable tools in machine learning. Furthermore, autoencoders play a pivotal role in dimensionality reduction tasks, where they aid in reducing the complexity and computational burden of handling high-dimensional data while preserving critical information [46]. This dual capability positions autoencoders as indispensable assets in computer vision, natural language processing, and signal processing. Ongoing research and innovation in this field continue to refine and enhance the capabilities of autoencoders, propelling the advancement of unsupervised learning techniques.

Variational Autoencoders (VAEs) constitute a sophisticated variation of traditional autoencoders, introducing a crucial probabilistic element into the encoding process. Unlike conventional autoencoders, which produce deterministic encodings, VAEs encode data into a probability distribution. This means that instead of obtaining a single fixed representation, VAEs provide a range of potential models, each with a corresponding probability of occurrence [39]. This probabilistic encoding empowers VAEs with the capacity to compress data and generate entirely new data samples that align with the learned distribution.

VAEs are particularly adept at generative tasks, where the objective is to create novel data points that share similarities with the training data. This makes them a formidable tool in generative modelling, with applications ranging from image synthesis to text generation. The ability to generate new data samples from a learned distribution has far-reaching implications, impacting fields such as computer graphics, natural language processing, and medical imaging, among others.

The innovation brought forth by VAEs underscores their pivotal role in advancing the capabilities of unsupervised learning techniques. Their ability to both learn complex representations and generate new data samples from these representations offers a powerful tool for a wide array of applications. Ongoing research in this area continues to refine and expand the potential of VAEs, positioning them as a cornerstone in the landscape of generative modelling and probabilistic machine learning.

Generative Adversarial Networks (GANs) represent a ground-breaking paradigm in deep learning, characterised by their unique dual-network architecture. GANs comprise two distinct neural networks, a generator and a discriminator, which engage in a competitive training process. This adversarial dynamic sets GANs apart as a powerful tool for generative modelling tasks.

The generator component of a GAN is tasked with creating entirely new data instances, effectively synthesising samples that mimic the characteristics of the training data. Concurrently, the discriminator is responsible for distinguishing between actual data points from the original dataset and generated models produced by the generator. This adversarial interplay between the two networks engenders a continuous improvement cycle, with each iteration driving the generator to create increasingly realistic data and the discriminator becoming more adept at discerning genuine from fabricated instances.

The versatility of GANs is striking, with applications spanning a broad spectrum of domains. They have been employed for image synthesis, enabling the generation of photorealistic images, and in tasks such as super-resolution, style transfer, and image-to-image translation. Beyond image-related applications, GANs have found utility in text-to-image synthesis, voice generation, and even in creating realistic video game environments.

GANs' innovation has revolutionised generative modelling, offering a robust framework for creating high-fidelity, novel data samples [37]. Ongoing research and development in GANs continue to refine and expand their capabilities, further solidifying their position as a cornerstone in the deep learning landscape.

Deep Belief Networks (DBNs) are a distinctive class of generative models characterised by their multi-layered architecture, comprising stochastic, latent variables [45]. This unique structure enables DBNs to excel in unsupervised learning tasks, mainly feature learning and dimensionality reduction.

At their core, DBNs are composed of multiple layers of hidden units, each interacting with the layer above it. This hierarchical arrangement empowers DBNs to capture intricate patterns and relationships within the data, making them particularly adept at tasks where understanding complex, high-level features is crucial.

The unsupervised learning capabilities of DBNs are precious in scenarios where labelled data is limited or unavailable. By leveraging the data's inherent structure, DBNs can autonomously discover meaningful representations, effectively reducing the dimensionality of the input space. This ability has profound implications in computer vision, natural language, and signal processing.

DBNs have demonstrated remarkable effectiveness in diverse applications, including but not limited to image recognition, speech analysis, and recommendation systems. Their adaptability and proficiency in unsupervised learning tasks make them a vital tool in the arsenal of machine learning practitioners [36]. Ongoing research and development in the field continue to refine and enhance the capabilities of DBNs, solidifying their significance in the landscape of deep learning models.

Initially conceived for Natural Language Processing (NLP) tasks, Transformers represent a ground-breaking architecture that leverages self-attention mechanisms to capture intricate global dependencies within data. This innovative approach revolutionised the field by allowing for parallelised processing of sequences, making them highly efficient for tasks requiring an understanding of long-range dependencies. Beyond NLP, Transformers have found extensive application in many functions, spanning language translation, text summarisation, sentiment analysis, and more. This versatility arises from their adaptability to tasks involving structured data where capturing relationships across distant elements is paramount.

One of the most influential derivatives of the Transformer architecture is Bidirectional Encoder Representations from Transformers (BERT), an acronym for Bidirectional Encoder Representations from Transformers. BERT is a pre-trained transformer model specially designed for natural language understanding tasks [35]. What sets BERT apart is its capacity to generate contextualised word representations, meaning it comprehends words based on their context within a sentence. This contextual awareness significantly enhances its ability to understand nuanced linguistic nuances, enabling it to excel in various NLP tasks, including sentiment analysis, named entity recognition, and question-answering.

Transformers and BERT have ushered in a new era in NLP, fundamentally transforming how machines comprehend and generate human language. Their influence extends across various applications, from automated customer support systems to content generation. Ongoing research and refinement in this domain continue to propel the capabilities of Transformers and BERT, pushing the boundaries of natural language understanding and age.

Capsule Networks (CapsNets) constitute a pioneering departure from the conventional Convolutional Neural Networks (CNNs), distinguished by their emphasis on capturing detailed information about the constituent parts of objects and the intricate spatial relationships between them [49]. This unique architectural approach marks a significant leap forward in object recognition, addressing a notable limitation of CNNs. Capsule Networks are particularly adept at understanding how different elements within an object interact and relate, making them exceptionally valuable in scenarios where discerning fine-grained details and handling viewpoint variations are critical.

The foundational concept behind Capsule Networks is the utilisation of capsules, specialised neural network units that encode information about specific features of an object and their relative positions. Unlike traditional neural networks that may struggle with transformations such as rotations or distortions, Capsule Networks have the potential to preserve the spatial relationships between object components, enabling them to handle viewpoint variations with greater efficacy.

Capsule Networks have shown remarkable promise in object recognition, pose estimation, and image reconstruction tasks. Their capacity to grasp the hierarchical structure of objects and their constituent parts has implications for fields as diverse as computer vision, robotics, and medical imaging [47]. The ongoing refinement and exploration of Capsule Networks continue to expand their potential, propelling them to the forefront of cutting-edge research in deep learning and computer vision.

Networks (CapsNets) constitute a pioneering departure from the conventional Convolutional Neural Networks (CNNs), distinguished by their emphasis on capturing detailed information about the constituent parts of objects and the intricate spatial relationships between them [46]. This unique architectural approach marks a significant leap forward in object recognition, addressing a notable limitation of CNNs. Capsule Networks are particularly adept at understanding how different elements within an object interact and relate, making them exceptionally valuable in scenarios where discerning fine-grained details and handling viewpoint variations are critical.

The foundational concept behind Capsule Networks is the utilisation of capsules, specialised neural network units that encode information about specific features of an object and their relative positions. Unlike traditional neural networks that may struggle with transformations such as rotations or distortions, Capsule Networks have the potential to preserve the spatial relationships between object components, enabling them to handle viewpoint variations with greater efficacy.

Capsule Networks have shown remarkable promise in object recognition, pose estimation, and image reconstruction tasks. Their capacity to grasp the hierarchical structure of objects and their constituent parts has implications for fields as diverse as computer vision, robotics, and medical imaging [39]. The ongoing refinement and exploration of Capsule Networks continue to expand their potential, propelling them to the forefront of cutting-edge research in deep learning and computer vision.

Neuroevolution of Augmenting Topologies (NEAT) is a pivotal advancement in artificial intelligence and neural network development. It represents an evolutionary algorithm meticulously designed to evolve artificial neural networks (ANNs) [50]. What sets NEAT apart is its capacity to dynamically adapt the structure and topology of neural networks throughout the evolutionary process.

NEAT employs a principled approach, introducing new neurons and connections over generations, allowing the network to grow in complexity and adapt to the evolving demands of the task. This unique methodology mitigates the common challenges associated with fixed-topology neural networks, such as finding the optimal architecture for a given problem.

The applications of NEAT are far-reaching, with a prominent focus on tasks involving reinforcement learning. By dynamically adjusting the network architecture and connections, NEAT enables the emergence of neural network structures tailored to the specific demands of complex, dynamic environments. This makes NEAT particularly powerful in scenarios where adaptability, robustness, and performance optimisation are paramount.

NEAT has had a transformative impact on artificial intelligence, with applications spanning robotics, game-playing, and control systems. Its adaptability and versatility have positioned NEAT as a foundational tool for researchers and practitioners seeking to harness the power of evolutionary algorithms in developing artificial neural networks [3]. Ongoing research and refinement in this domain continue to expand the potential of NEAT, driving advancements in neuroevolution and adaptive learning.

This comprehensive review paper explores the prevalent deep learning architectures utilised in animal detection, focusing on methods grounded in Convolutional Neural Networks (CNNs). By examining and analysing the application of CNN-based techniques, this review endeavours to provide a thorough understanding of their effectiveness and potential in advancing the field of animal detection [37]. Through an in-depth exploration of the various CNN-based approaches, this paper aims to shed light on the state-of-the-art methodologies and their contributions to enhancing animal detection systems' accuracy and efficiency.

A. *Convolutional Neural Networks (CNNs)*

Convolutional Neural Networks (CNNs) are deep learning methods well-suited for image-based tasks. They are designed to mimic the human brain's visual processing and are collected from multiple layers of convolutional filters and pooling operations [22]. These layers allow the CNN to learn hierarchical features from the input images, enabling the detection of complex patterns and objects.

Convolutional Neural Networks (CNNs) are a specific feed-forward artificial neural network type. Their structural organisation is influenced by the arrangement of cells in the animal visual cortex. Within the visual cortex, small clusters of cells exhibit sensitivity to specific areas of the visual field. Neuronal cells in the brain respond selectively, firing only in the presence of edge orientations. For instance, some neurons activate in the fact of vertical edges, while others do so for horizontal or diagonal edges. CNNs, employed in deep learning, are designed to assess visual information [49]. They can tackle a wide array of tasks, including processing images, sounds, texts, videos, and various other forms of media.

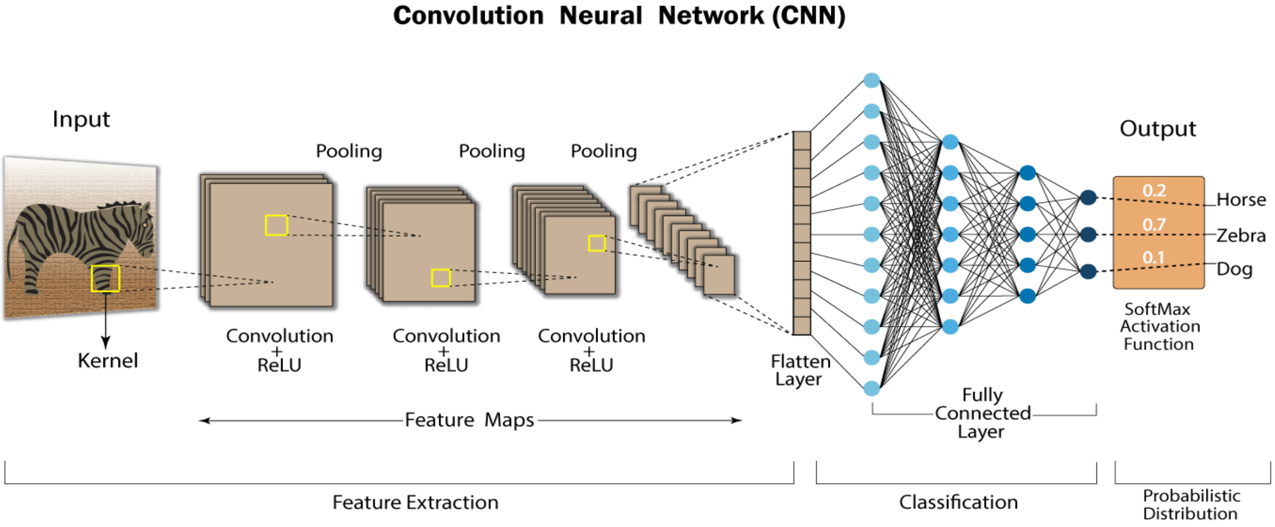

Fig. 3. FA Graphical representation of Convolutional Neural Network (CNN).

Convolutional Neural Networks (CNNs) have an input layer, multiple hidden layers, an output layer, and many parameters, enabling them to discern complex objects and intricate patterns adeptly. These networks employ convolution and pooling processes to down-sample the input data before applying an activation function. These operations are predominantly carried out within partially connected hidden layers, culminating in a fully connected layer that yields the output shown in Fig. 3.

The resultant output from a CNN maintains a spatial dimensionality like the original input image Convolution, in this context, involves the amalgamation of two functions to generate the output of the latter function. In CNNs, the input image undergoes convolution by applying filters, yielding a Feature map. These filters comprise randomly generated vectors encompassing weights and biases within the network [42]. Unlike individualised weights and preferences for each neuron, CNNs employ uniform weights and biases across all neurons. Multiple filters can be instantiated, capturing distinct facets from the input data. Filters are alternatively referred to as kernels.

In animal detection on video data, CNNs play a crucial role in feature extraction and recognition. The initial layers of a CNN identify modest features with similar edges and textures, while deeper layers learn huge abstract and discriminative features, which are relevant for detecting animals. CNNs can be fine-tuned with labelled animal images to adapt the model for animal detection tasks. Convolutional Neural Networks (CNNs) are extensively used, including variants like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), R-CNN, Mask R-CNN and RetinaNet, which offer real-time or high-accuracy object detection. Temporal aspects are addressed by I3D (Inflated 3D ConvNet) and T-CNN (Temporal CNN), which capture motion cues. For instance, Mask R-CNN and Tube-CNN are applied for segmentation and tracking. Spatio-temporal interactions are emphasised by STSN (Spatio-Temporal Snippet Network) and STAM (Spatio-Temporal Attention Mechanism). More recent approaches utilise boundary matching (BMN) and trajectory forecasting (TrajNet++) for improved tracking [41].

These architectures leverage deep learning's capacity to learn intricate patterns and temporal dependencies, enabling accurate and efficient animal detection in video data.

1) Region-based CNNs: Region-based Convolutional Neural Networks (CNNs) are deep learning models specifically designed for tasks like object detection. These architectures accurately and efficiently detect objects within images or videos [30]. Within the realm of RCNNs, several algorithms have been developed to improve the performance and efficiency of animal detection. Some of these include:

Faster R-CNN combines region proposal networks (RPN) with a CNN-based object detector. The RPN proposes candidate regions likely to contain animals, and then the CNN refines and classifies these regions. This two-stage approach improves detection accuracy by focusing on promising areas rather than scanning the entire image. R-FCN (Region-based Fully Convolutional Networks) [48] is a single-stage detector that operates directly on the whole picture, using position-sensitive score maps to predict object locations. It achieves accurate animal detection while being computationally efficient.

Fast R-CNN, a pioneering development in object detection, ushered in a paradigm shift by introducing a unified framework that seamlessly integrates region proposal generation and object classification within a single pass. This streamlined approach accelerates detection and enhances accuracy [46]. One of its key innovations is the integration of a Region of Interest (RoI) pooling layer, which efficiently extracts features from various regions of an image. This optimises computational resources and facilitates handling multiple areas of interest.

The impact of Fast R-CNN was profound, transcending previous approaches in both speed and precision. This breakthrough paved the way for a new generation of object detection systems, setting a high bar for subsequent advancements in the field [19]. Fast R-CNN's efficiency and accuracy have made it an instrumental tool in diverse applications, from computer vision tasks like image recognition and scene understanding to practical applications in fields such as autonomous vehicles, surveillance systems, and more. Its influence continues to resonate in the ongoing evolution of object detection methodologies.

Faster R-CNN stands as a remarkable refinement in object detection, building upon the foundation laid by Fast R-CNN. This innovative architecture introduced a pivotal component known as the Region Proposal Network (RPN). Unlike its predecessors, which relied on external algorithms for region proposal generation, the RPN operates directly on the feature maps [41]. This streamlined approach eliminates additional computations, significantly enhancing computational efficiency.

By seamlessly integrating the RPN, Faster R-CNN achieves a remarkable fusion of region proposal generation and object classification within a unified framework. This integration expedites the detection process and leads to substantial gains in accuracy. The direct generation of region proposals from the feature maps represents a significant leap forward, enabling Faster R-CNN to outperform its predecessors in speed and precision.

Faster R-CNN's introduction of the RPN revolutionised the field of object detection, establishing it as a cornerstone in modern computer vision. Its impact extends across a wide array of applications, including but not limited to autonomous driving, object tracking, and facial recognition. The efficiency and accuracy achieved by Faster R-CNN have solidified its status as a foundational framework in the landscape of object detection methodologies [27]. The ongoing research and development in this area continue to build upon the innovations by Faster R-CNN, further advancing the capabilities of object detection systems.

Mask R-CNN represents a monumental advancement in object detection that builds upon the formidable Faster R-CNN framework. What sets Mask R-CNN apart is incorporating a third critical branch dedicated to predicting object masks, bounding boxes, and class probabilities. This breakthrough innovation introduces a level of granularity that was previously unparalleled.

Mask R-CNN achieves a monumental leap forward in object understanding by enabling the precise delineation of objects within an image [17]. This capability, known as instance segmentation, has wide-ranging applications in tasks where detailed object comprehension is paramount. It allows for accurately identifying objects and differentiating between individual instances of the same class.

The addition of the mask prediction branch in Mask R-CNN has revolutionised the field of computer vision and object detection. It has found extensive use in domains such as medical imaging, robotics, and autonomous navigation, where discerning detailed object boundaries are crucial. Mask R-CNN has solidified its position as an indispensable tool for tasks demanding high precision in object localisation and segmentation.

The ground-breaking contributions of Mask R-CNN continue to resonate in computer vision, inspiring further innovations and advancements in object detection and instance segmentation techniques. Its impact extends across a broad spectrum of industries and applications, showcasing its pivotal role in advancing the capabilities of visual perception systems.

Cascade R-CNN, a significant evolution in object detection, introduces a multi-stage approach to enhance object proposal refinement. This innovative architecture deploys a succession of classifiers with progressively higher difficulty thresholds to filter out potential false positives systematically. This cascade strategy fundamentally improves precision in object detection, setting Cascade R-CNN apart as a crucial model for tasks where pinpoint accuracy is paramount.

By employing a cascade of classifiers, Cascade R-CNN effectively refines the object proposal process through stages of increasing stringency. This meticulous filtering mechanism substantially elevates the model's ability to discriminate between true positive and false positive detections. As a result, Cascade R-CNN achieves a level of precision that surpasses previous object detection models.

This refined approach has widespread applications in medical imaging, robotics, and aerial imagery analysis, where high detection accuracy is critical. The cascade architecture in Cascade R-CNN provides a powerful tool for tasks that demand meticulous object recognition, making it an indispensable asset in the arsenal of computer vision practitioners.

The introduction of Cascade R-CNN exemplifies the ongoing pursuit of precision and accuracy in object detection methodologies. Its impact reverberates across a broad spectrum of industries and applications, showcasing its pivotal role in advancing the capabilities of visual perception systems. The continued refinement and exploration of Cascade R-CNN continue to drive progress in object detection.

Indeed, Faster R-CNN and Mask R-CNN are well-suited for animal detection tasks necessitating precise localisation, as they precisely identify the boundaries of objects within an image. Moreover, their adaptability extends to video-based applications, allowing real-time or near-real-time animal detection in dynamic environments.

Region-based Convolutional Neural Networks (CNNs) like Faster R-CNN and Mask R-CNN strike a vital balance between accuracy and computational efficiency. This characteristic makes them highly versatile and applicable in various animal detection scenarios whether in wildlife monitoring, conservation efforts, or ecological research, these models offer a robust solution for accurately identifying and localising animals within imagery or video footage.

The adaptability and efficacy of Faster R-CNN and Mask R-CNN have solidified them as foundational tools in computer vision for animal detection. Their versatile application extends across various domains, showcasing their pivotal role in advancing the study of understanding and animal detection.

2) Single Shot Multibox Detector (SSD): SSD is a real-world object detection algorithm in the one-stage detectors category. It is widely used for animal detection due to its efficiency and accuracy. SSD works by separating the original image into multiple grids, and every grid cell is answerable for forecasting bounding boxes and class probabilities for potential objects. These predictions are made at various scales, viewing the model to handle varying sizes.

One of the main advantages of SSD is its speed and suitability for real-time animal detection in video streams. It can process video frames rapidly, making it ideal for applications where low latency is essential. Moreover, SSD can efficiently detect multiple animal instances in a single forward pass, making it highly scalable for large-scale animal monitoring scenarios [16].

3) You Only Look Once (YOLO): YOLO is an alternative real-time object detection architecture recognised for its speed and simplicity. Unlike SSD, YOLO approaches object finding as a regression problem. It divides the input image into a grid, and every grid cell directly predicts bounding box coordinates and class probabilities without using separate anchor boxes.

YOLO's characteristics make it a strong candidate for video-based animal detection. Its single-pass architecture enables real-time processing of video frames, making it well-suited for applications that require low latency. However, YOLO might struggle with detecting small animals or closely grouped instances due to the anchor boxes, which could affect its localization [12]. Deep learning architectures have revolutionised animal detection in computer vision. Convolutional Neural Networks provide a solid foundation for feature extraction, while SSD and YOLO offer real-time capabilities, making them ideal for video-based animal detection.

Various authors have applied deep learning to animal detection. In a [1] study, a Convolutional Neural Network (CNN) algorithm was employed to achieve 93.8% accuracy in identifying 48 species in the Snapshot Serengeti dataset, offering automated animal identification with high precision. However, this approach didn't utilise pre-processing techniques or filters. In another [4] study, tracking and detection models like BYTETrack, SORT, IoU-tracker, YOLOv4, DeepSORT, and Few-MOT were used to monitor endangered animals, recording their daily movements and activity areas. While this method embedded uncertainty into multi-object tracking for robust models, it relied on limited frames for experimentation.

In a [5] study, various deep neural networks, including AlexNet, NiN, VGG, GoogLeNet, ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 were employed, achieving an impressive 96.8% accuracy in identifying animals in camera-trap images. This method successfully identified, counted, and described animals using deep neural networks, although it was limited to a specific dataset (SS dataset). In study [6], the study focused on utilising convolutional neural networks, emphasising day-night joint training and YOLOv5, resulting in a high accuracy of 97.9%. While this approach demonstrated high accuracy, it was considered time-consuming and labour-intensive, mainly due to the difficulty in handling small datasets. Additionally, a [7] study utilised Cascade R-CNN, HRNet32, ResNet50, and ResNet101, achieving a performance of 97%, with the added advantage of efficient imagery processing and saving time. However, it was observed that R-CNN occasionally generated incorrect candidate region proposals. Lastly, in [8], a method combining HOG/SVM and deep neural networks (Faster RCNN and YOLO) attained an accuracy of 87%. This approach excelled in generalising images from the web but faced challenges in detecting small objects.

These studies, among others, highlights the diverse applications of deep learning in animal detection, achieving high accuracy rates and automation benefits but facing challenges in data requirements, computational intensity, and generalisation. The choice of architecture depends on specific application requirements, such as real-time performance, detection accuracy, and computational resources, as research in deep learning continues advancing more effective and efficient architectures for animal detection.

III. The Process of Animal Detection Process from Video Data

Animal detection refers to identifying and recognising animals' presence or location in a given environment. This can be done through various means, including visual observations, sensor technologies, or automated systems utilising computer algorithms.

In technology and computer science, animal detection often involves using machine learning or deep learning techniques to analyse images, videos, or sensor data to identify and classify animals. This technology is used in various applications, including wildlife conservation, agriculture, surveillance, and research. For example, animal detection systems may be deployed in wildlife conservation to monitor the movement and behaviour of endangered species. In agriculture, such systems can track livestock or identify pests. In research, animal detection technology helps gather animal behaviour and ecology data.

Overall, Animal detection is crucial in advancing our understanding and management of animal populations and various industries that interact with or rely on animals. Using low-cost commercial drones, artificial intelligence (AI), neural networks, and computational power has simplified detecting items of interest. Deep learning and convolutional neural networks (CNNs) algorithms are now the benchmark in picture-processing jobs such as object recognition and segmentation [5]. These networks are critical instruments for identifying and analysing animals in video recordings. "Graphical Processing Units (GPUs)" are now widely used in the digital vision field, especially those seeking deep learning and lowering model learning and inference time. Many contemporary conservation methods use machine learning to analyse pictures after data gathering. When images are gathered, machine learning must be implemented.

A. Data Collection from Different Datasets

In study [5], only one data collection was created and used. This included seven hundred photos. The Black had Red, Green, Blue and black/white images. The information was divided into two categories: rhinoceros and automobiles. Each class had 350 picture sizes ranging from 300 *147 to 3840* 2160 pixels. The data was divided into three distinct datasets: one for instruction, another for confirmation, and one for testing. Observing that the testing sample is not used during the model's training is essential. It is exposed in Fig. 4.

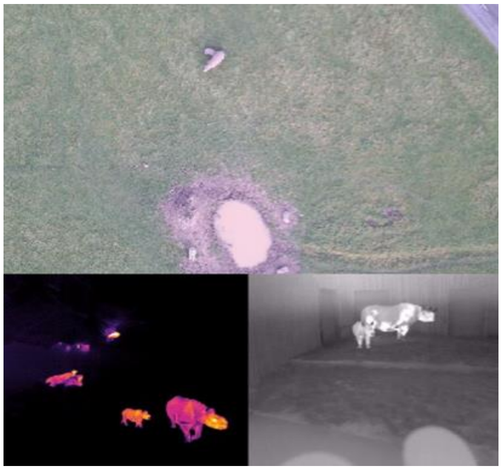

Fig. 4. Sample of training data with variations.

In study [6], the video segments used were captured by infrared detectors in the "Northeast Tiger and Leopard National Park between 2014 and 2020". It chose 17 significant types. It extracted pictures from videos using a "Python script" at a frame interval of fifty. It manually consistently annotates the pictures. It is exposed in Fig. 5.

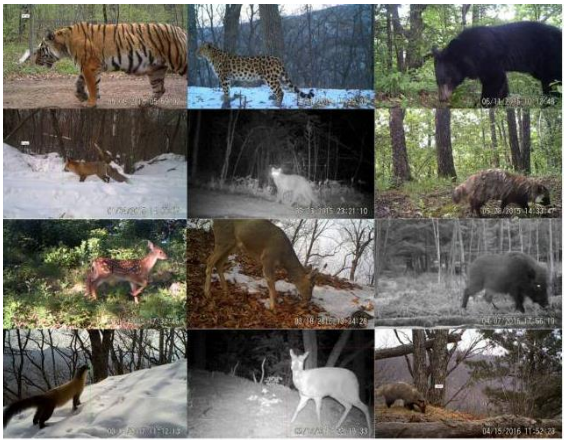

Fig. 5. Sample of the dataset with some species.

In research [7], a dataset with thermal pictures of two mammal groups was created: red deer, European roe deer, and fallow deer and swine, which primarily comprised images of European wild boar. The sample was created using a Pulsar Helion 2 XP50 PRO camera.

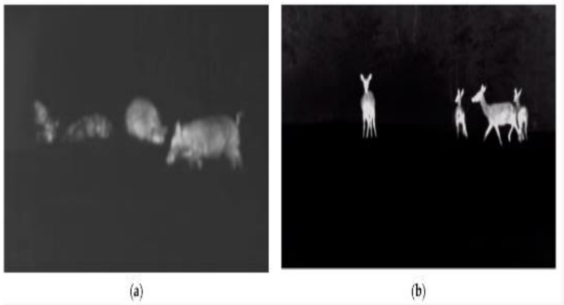

Fig. 6. Example of two distinct objects: (a) "wild boar" and (b) "deer."

One of the most important and widely recognised datasets in animal detection in video data is the "ImageNet Large Scale Visual Recognition Challenge" (ILSVRC). Although primarily focused on object recognition in images, ILSVRC includes a subset of classes related to animals. This dataset was pivotal in advancing deep learning techniques for object detection, serving as a benchmark for various computer vision tasks. In the domain of animal detection specifically, the "ADE20K" dataset is significant. It is a diverse dataset that includes images with annotations for object detection, semantic segmentation, and scene parsing, as shown in Fig. 6. While not exclusively dedicated to animals, it provides a valuable resource for researchers working on animal detection in complex visual environments.

Another prominent dataset is the "COCO (Common Objects in Context)" dataset, widely used for object detection and segmentation tasks. It contains many annotated images depicting a broad range of objects in complex scenes, including various animals in diverse contexts. For more specific applications, datasets like the "AI4MARS - Annotated Image for Machine Learning in Animal Recognition System" focus on animal detection in particular environments, such as the Mars Rover mission. This dataset is curated to detect and classify animals in Martian terrain.

Regarding benchmark datasets, the "PASCAL Visual Object Classes" (PASCAL VOC) dataset is renowned. It covers many object classes, including animals, and has been extensively used for evaluating object detection algorithms. It provides a standardised evaluation platform for researchers in the computer vision community.

The "MS COCO" dataset, in addition to being a significant dataset, also serves as a benchmark for object detection tasks. It includes various object categories, including animals, and is accompanied by a comprehensive evaluation metric suite. These datasets and models play a crucial role in advancing the field of animal detection in video data. They provide standardised and diverse sets of images with ground truth annotations, enabling researchers to train and evaluate their algorithms consistently. This ensures that advancements in animal detection techniques are rigorously tested and compared against state-of-the-art methods, ultimately driving progress in the field.

Some of the datasets from recent works are listed in Table II below:

TABLE II. DETAILS OF DATASET FOR ANIMAL DETECTION VIDEO USING DEEP LEARNING FROM RECENT WORKS

References	Datasets	Description	Scenario
[5, 2018]	ImageNet	It has 1.3 million labelled images for 1,000 groups (from synthetic objects, for example, bicycles and cars, to wildlife categories 1q11 dogs and lions)	Identify and count wild animals in the camera trail camera.
[30, 2021]	Bavarian Highway Directorate, Germany	It displays video segments that last around 10 seconds and has an eight fps (frames per second) resolution of 1280 720 pixels. The footage was captured by camera trail cameras set up at an "animals' bridge" (wildlife crossing) on Federal Highway 7 near "Oberthulba."	To identify animals in wildlife videos
[31, 2018]	Fly and mouse datasets	59 aligned, high-resolution behavioural videos	To estimate the fast animal pose.
[32, 2023]	PolarBearVidID	It includes video sequences of 13 polar bears in various poses and lighting conditions..	Identify animal behaviour in zoos.
[33, 2020]	coco dataset	It has a width and height in the range of 40 and 140 pixels	To detect, classify, and track animals in the African savannah
[34, 2019]	Badger dataset	Images were captured at a selection of UK farms where surveillance occurred. All were manually assigned to badger, bird, cat, fox, rat, or rabbit.	To monitor wildlife

B. Pre-processing

Data pre-processing is critical to get the best results from any artificial intelligence-based approach. Various data pre-processing methods can be deployed to make the information fit for the development of models. If the data taken from the devices was noisy and meaningless, it was deleted from the collection. The research in [8] used a "6th-order Butterworth filter with a cutoff frequency of 3.667 Hz" to eliminate the noise and anomalies from the information. In [9], YOLOv5 eliminates duplicate and similar images. In [10], by employing computer vision tools, a picture processing system was created and built to enhance contrast in pictures and segment pertinent image subdivisions. The photos were revised to expedite the procedure. It is shown in Fig. 6.

C. Image Segmentation

Image segmentation involves partitioning each video frame into distinct regions corresponding to specific objects or animals. This enables the isolation and identification of animals within a dynamic visual stream. Techniques like thresholding, edge detection, and deep learning-based segmentation are employed to delineate animals from the background, allowing for subsequent analysis such as tracking, behaviour monitoring, and species classification throughout the video sequence. Effective segmentation is crucial for accurate and reliable animal detection and tracking in dynamic video data. A computer vision method locates items within an image by forming a bounding box surrounding it to comprehend what is in it at the pixel stage. In study [7], the bear was segmented from the picture using the "MSER (Maximally Stable Extremal Region)" method. In study [11], it uses Mask R-CNN for segmentation. It can distinguish between different items in a video. It returns the class identity, item masks, and the bounding box coordinates for every object in a provided picture. The process is split into two phases; the first scans the image and generates a "Region proposal network (RPN)" to provide potential object bounding boxes, and the second differentiates proposals and generates bounding boxes and boundaries for every class. The study in [12] used the "Falzenszwalb algorithm" for image segmentation, which organises pixels with comparable luminance.

D. Classification

In animal detection, classification pertains to categorising identified objects into specific animal species or classes based on distinctive features learned through the model. It distinguishes different animals within a scene, aiding in species identification and population monitoring. [30] In animal detection, classification is crucial for understanding wildlife dynamics, ecological studies, and conservation efforts shown in the Fig. 7. It creates a list of desired objects for Detection images and instructs a model to identify them using labelled sample photos. In [5], Faster-RCNN is used for image classification in video clips. In [6], "Dilated Residual Networks (DRN)" was used, and better accuracy was achieved. In [12], the author used deep boosting, dictionary learning, and convolution networks for image classification. In [9], three models ("YOLOv5m, CNN_HRNet32, FCOS_Resnet101") were used for simulation to detect and classify multiple animals in a video, and the video classification accuracy of FCOS_Resnet101 achieved 91.6%. It is shown in Fig. 5.

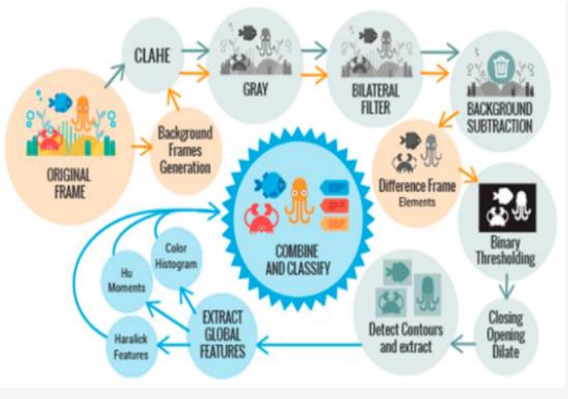

Fig. 7. Pipeline for data pre-processing.

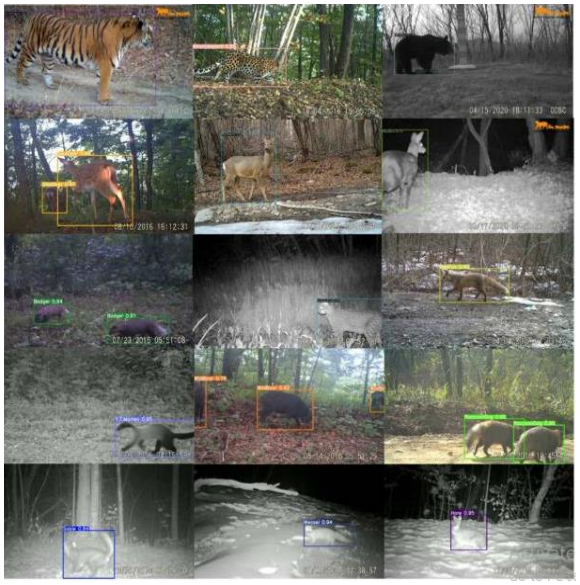

Fig. 8. Example of correct classification.

E. Evaluation Metrics for Animal Detection for Video Data

The evaluation metrics for animal detection in video data include precision, recall, F1 Score, mean average precision (mAP), intersection over union (IoU), false positive rate (FPR), false negative rate (FNR), accuracy, mean recall, processing speed, and frame-level vs. video-level evaluation [13]. These metrics assess the system's ability to accurately detect animals, handle false positives and negatives, and its overall performance in real-time video processing, providing a comprehensive evaluation of the detection system's effectiveness and efficiency. Fig. 8 shows example of correct classification.

1) Video data evaluation metrics: Evaluating video-based animal detection systems is vital to measure their accuracy and effectiveness in real-world scenarios. As deep learning models are commonly used for this task, specific evaluation metrics are required to assess their performance on video data [14]. One of the key metrics used in this context is the mean average precision at different Intersections over Union (mAP@IoU) thresholds over time.

mAP@IoU is an extension of the traditional mean average precision (mAP) metric, commonly used in image object detection tasks. It measures the precision and recall of predicted bounding boxes by comparing their overlaps with ground-truth annotations at different IoU thresholds. This provides a comprehensive evaluation of the model's ability to accurately detect animals in video sequences across various IoU thresholds, reflecting different levels of object localisation accuracy [16]. Evaluating mAP@IoU over time enables the assessment of the model's consistency and robustness in detecting animals throughout the video, accounting for spatial and temporal variations.

2) Temporal consistency metrics: Temporal consistency is a critical aspect of video-based animal detection as it ensures stable and reliable tracking of animals across frames. Several methods can be employed to evaluate the stability and consistency of animal detections over time:

Temporal Intersection over Union (tIoU): tIoU measures the temporal overlap between predicted bounding boxes and ground-truth annotations in consecutive frames [17]. It quantifies how well the model can maintain accurate and continuous detections of animals across time. A higher tIoU score indicates better temporal consistency and reliable tracking.

ID Switching Rate: The ID switching rate evaluates how often the model incorrectly assigns different identities to the same animal or switches identities between consecutive frames [28]. A lower ID switching rate signifies improved temporal consistency in tracking and maintaining individual animal identities.

Fragmentation Rate: Fragmentation occurs when the model fails to link consecutive detections of the same animal, resulting in disjointed tracks [15]. The fragmentation rate measures the degree of this issue, with lower values indicating better temporal consistency and continuous tracking.

Trajectory Smoothness: This metric assesses the smoothness and continuity of animal trajectories over time. [18] A model with high trajectory smoothness exhibits more consistent and visually coherent animal tracks, indicating reliable temporal consistency.

By integrating video-based evaluation metrics like mAP@IoU over time with temporal consistency metrics, researchers and practitioners can thoroughly assess the performance of deep learning models for video-based animal detection [22]. These evaluation measures help identify potential areas for improvement, validate the suitability of the models for specific applications like wildlife monitoring, behaviour analysis, and ecological research, and drive advancements in computer vision for animal-related studies.

IV. Recent Progress in Animal Detection using Deep Learning

Recent Progress in animal detection using different machine learning and deep learning techniques is presented here. Table III, in a tabular format, compares the various DL and ML learning models for animal detection. A comparison is made regarding the detection methods, main findings, advantages, and disadvantages for further research from existing literary works.

These references represent a range of scholarly works that have explored various deep-learning techniques applied explicitly to video-based animal detection. [19] By including these references, the review aims to provide a comprehensive and up-to-date analysis of the existing literature in this domain. [33] Additionally, these selected references contribute significant insights, methodologies, and findings relevant to understanding the advancements, challenges, and potential applications of deep learning in video-based animal detection. [27] Their inclusion strengthens the credibility and rigour of the review, offering a solid foundation for examining the state-of-the-art approaches and identifying future research directions in this field.

TABLE III. RECENT PROGRESS IN ANIMAL DETECTION USING DIFFERENT MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

References	Detection Methods	Main Finding	Advantages	Disadvantages
[1, 2020]	Convolutional Neural Network (CNN) algorithm to detect wild animals.	Deep convolutional neural networks to identify, count, and describe the behaviours of 48 species in the 3.2-million-image Snapshot Serengeti dataset. 93.8% accuracy	Automating animal identification with 99.3% accuracy, matching human volunteers' 96.6% saves over 17,000 hours of labelling effort on a 3.2-million-image dataset.	No Pre-processing techniques or filters are used in this paper.
[4, 2022]	BYTETrack, SORT, IoU-tracker, V-IoU-tracker, YOLOv4, + DeepSORT and Few-MOT model	Few-MOT for wildlife to embed uncertainty into designing a multiobject-tracking model by combining the richness of deep neural networks with few-shot learning, leading to correctable and robust models.	Few-MOT employs few-shot learning and tracking-by-detection to monitor endangered animals, recording daily movements and frequent activity areas for analysis.	A limited number of frames are used to experiment.
[5, 2018]	AlexNet, NiN, **VGG,** GoogLeNe, ResNet-18, ResNet-34, ResNet-50, ResNet-101 , ResNet-152	VGG model achieved the best accuracy of 96.8%	Deep neural networks (DNNs) can successfully identify, count, and describe animals in camera-trap images.	Only an SS data set is used.
[6, 2021]	The detection of animals Using convolutional neural networks.	Day-night joint training had a better performance.YOLOv5 achieved an accuracy of 97.9%	Time-consuming and labour intensive.	Difficulty with small datasets
[7, 2022]	Cascade R-CNN, HRNet32, ResNet50 and ResNet101.	It achieved a performance of 97%.	The imagery potentially quickly and efficiently saves much time.	R-CNN sometimes generates wrong candidate region proposals as the selective search is a fixed algorithm with no learning capabilities.
[8, 2023]	(HOG/SVM), deep neural networks (Faster RCNN and YOLO).	It attained 87% accuracy	The web can generalise the image better.	Struggles to detect small objects.
[9, 2022]	Convolutional neural networks (CNNs) 1D CNN and 2D CNNs	It attained 99.70% training accuracy and 96.85% validation accuracy.	Real-time monitoring of activities	High computational requirements
[10, 2022]	Deep learning technology	Accuracy of 95.1%	It can benefit from computer automatic identification of postural behaviour, which can be used to quantify animal activity.	This can be costly and time-consuming.
[11, 2020]	Machine learning and Deep learning.	VGGNET was the best algorithm for an accuracy of 96.6 %.	Monitoring networks capable of providing large amounts	Movies cannot all be manually processed
[12, 2022]	YOLO; Convolutional neural networks (CNN); SSD; mask R-CNN; VGG-Net	VGGNET was the best algorithm for animal classification, with an accuracy of 96.6 %	It has a standard and easy-to-understand architecture for CNNs, with multiple convolution and pooling layers.	Slow training time
[13, 2023]	Deep learning-based model for automated recognition, definition, and numbering of wild creatures in camera trail camera images	A deep learning model automates the recognition, classification, and counting of wildlife in trail camera images from the "Snapshot Serengeti dataset.	This deep learning model automates animal recognition and description in trail camera images, streamlining manual work and scaling for large datasets.	Data quality impacts accuracy: biased or poorly labelled data may yield suboptimal results. Limited generalisation, computational demands, and interpretability are challenges.
[14, 2019]	Evaluation of ML and DL techniques, including SVM, RF, AlexNet, and Inception v3, for classifying animal genera from the "KTH dataset.	ML and DL techniques for Detection of animal genera using camera capture images. They investigated "SVM, RF, deep learning methods, and AlexNet and Inception v3 machine learning techniques". They used the "KTH dataset," which includes nineteen distinct animal categories.	The paper compares ML and DL techniques for animal genera recognition, emphasising DL's automatic feature learning, scalability, and generalisation.	ML and DL techniques depend on diverse data for better performance. Hyperparameter tuning is time-consuming. DL requires more data, which is challenging for specific animal genera. Interpretability is difficult, especially in complex, deep architectures.
[15, 2019]	Random forest, K-nearest neighbours (KNN), support vector machine (SVM), naive Bayes, and artificial neural network (ANN)	The study showcased a computerised pet movement and mood detection system, exploring data sources and machine learning techniques like random forest, KNN, SVM, naive Bayes, and ANN.	Random Forest for complex relationships, KNN for non-linear data, SVM for high dimensions, Naive Bayes for text/categorical data, and ANN for versatile deep learning applications.	ML algorithms have unique strengths: Random Forest for complex relationships, KNN for non-linearity, SVM for high dimensions, Naive Bayes for text/categorical data, and ANN for versatile deep learning.
[16, 2018]	The authors employ two popular object detection methods - Single Shot Multi-Box Detector (SSD) and You Only Look Once (YOLO).	Set forward a paradigm for automatically identifying animals via camera-trail camera ped pictures. It focused on determining the species and the number of species recorded in	SSD and YOLO enable real-time object detection, detecting multiple objects in one pass. Their end-to-end approach is versatile for animal species without retraining.	Training deep learning-based object detection models like SSD and YOLO demands a large, labelled dataset that is resource-intensive to collect.

		the photograph through algorithms for object detection like "Single Shot Multi-Box Detector (SSD) and You Only Look Once (YOLO).”		Performance varies with image complexity, requiring hyperparameter tuning. False positives/negatives affect accuracy.
[17, 2020]	We identify bird species using a "Super-resolution Mask RCNN-based transfer deep learning" model. The method combines super-resolution and Mask RCNN techniques to identify bird species accurately.	"Super-resolution Mask RCNN-based transfer deep learning approach" to identify bird species. They used Mask RCNN to identify very minute differences that analysed pixel-by-pixel of the picture and added a mask to it, making it easy to identify the dimensions and form of the item.	Combined with super-resolution, Mask R-CNN improves bird species identification, providing precise localisation and reducing overhead through transfer learning.	Mask RCNN and super-resolution demand significant computational resources. Transfer learning reduces data needs, but diverse bird datasets are vital for fine-tuning. Mask RCNN lacks interpretability.
[18, 2018]	Recent deep-learning methods are used to surmount the expense and effort of processing camera trail camera pictures.	The authors employ two distinct datasets to monitor populations of animals and govern environments all over the globe.	Deep learning automates camera trap image analysis, aiding animal identification, counting, and achieving state-of-the-art accuracy for global monitoring.	Deep learning requires ample labelled data, which is challenging for rare species. High-performance hardware is vital due to computational intensity. Model interpretability and bias are complex.
[19, 2020]	The proposed technique for animal identification employs neural network design such as "SSD and faster R-CNN.”	Using object identification, an exclusive animal recognition and collision avoidance system aims to detect animals and prevent road collisions.	SSD and Faster R-CNN offer real-time object detection, including animals, for collision avoidance systems, enhancing road safety for humans and wildlife.	Deep learning object detection models like SSD and Faster R-CNN rely on expensive labelled datasets susceptible to environmental variations. Generalisation hinges on diverse data, with the potential for false positives/negatives and performance influenced by model settings and training data.
[20, 2021]	The suggested technique is based on the "Sobel edge algorithm," which is basic but effective in detecting edges based on modified values.	The method achieves rapid detection (0.033 s per image) through thermal pixel analysis, synchronising thermal and RGB attributes for superior real-time animal detection across diverse settings. "Size-temperature filters" enhance applicability.	The real-time system aids wildlife monitoring, safety, and research sans extensive training data, with contextual information reducing false positives bolstered by thermal and RGB fusion.	The Sobel edge algorithm's limitation lies in detecting only edges, lacking fine-grained details and colour information for precise species identification. Environmental conditions impact detection accuracy, and species-specific recognition is limited.
[21, 2020]	It evaluates performance forecasting incrementally ahead with long-range scenarios using "Random Forests, Neural, and Recurrent Neural Networks.” This method is used to analyse excellent quality and movement statistics. For one step forward prediction, it was discovered that individual-level Machine Learning and Deep Learning approaches beat the SDE model.	The broad framework for forecasting animal movement consists of two steps: initially estimating behavioural motion stages and then predicting the animal's velocity. This framework is specified for both individual and group training.	The framework assesses individual and group training for complete animal movement patterns, utilising Random Forests, Neural Networks, and Recurrent Neural Networks for varied prediction approaches. High-quality data enhances precision in predictions.	Effectiveness relies on adequate data, which could be more challenging. Multiple models add complexity, demanding computational resources. Performance varies with species and environment, affecting generalisation.
[22, 2021]	The DL algorithm detects and segregates the animals with a large-scale publicly available dataset. A CNN model predicts the object (animal) in every image frame obtained from the live camera.	They have implemented deep learning for detecting the animals from the videos to improve safety. If the algorithm detects the object as an animal, the system will generate an alarm for 3 seconds to avoid a collision. Results show that the proposed approach achieves an accuracy of 91%.	Achieving 91% real-time Animal detection accuracy, the approach uses deep learning and CNNs with large-scale dataset training for enhanced safety via collision alerts across diverse species and conditions.	The DL algorithm's accuracy and generalisation depend on the training dataset's quality and diversity. Environmental conditions impact accuracy. Detecting rare species poses challenges. False positives/negatives may occur.
[23, 2017]	A deep CNN model is employed in this work for animal detection, and the model is trained using a single labelled dataset. The DCNN algorithm can automatically filter animal images and identify animal	The proposed approach achieves a phenomenal accuracy of 96.6% for animal detection and 90.4% for identifying the species accurately.	The method achieves an impressive 96.6% accuracy in animal detection and 90.4% in species identification, which is vital for reliable wildlife monitoring.	Single-dataset reliance hampers generalisation to new environments and species, influenced by data bias. Diverse dataset training enhances performance but demands substantial computational resources, which is challenging

	species.			on resource-constrained platforms.
[24, 2017]	CNN model for animal recognition is implemented. The effectiveness of different image recognition techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Local Binary Patterns Histograms (LBPH), and Support Vector Machine (SVM) are analysed, and the performance is compared with the proposed CNN model in terms of recognition rate.	The models are experimentally evaluated for recognition time using a 500-set, 100-image animal database. PCA outperforms LDA and LBPH for large databases, while LBPH excels with small datasets. The proposed CNN model achieves 98% recognition accuracy compared to other models.	The CNN model achieved 98% accuracy in animal recognition, excelling in image identification with automatic feature learning, highlighting its superiority over traditional methods in animal recognition tasks.	CNN model performance depends on dataset quality and size, influenced by biases. It's resource-intensive, with hyperparameter tuning challenges, and offers limited interpretability as a black-box model.
[25, 2021]	The CNN model groups animal images based on the input database. The performance of CNN is compared with conventional recognition techniques such as SU, DS, MDF, LEGS, DRFI, MR, and GC. These techniques are usually characterised by high false positive and negative rate detection.	Developing an efficient animal recognition system is vital. Genetic algorithm-based image segmentation and neural network classification improve accuracy. A dataset with 100 subjects and two classes demonstrates higher precision (99.02%), recall (98.79%), F1 score (98.9%), and low MAE (0.78%).	A CNN model with genetic algorithm segmentation achieves high animal image accuracy, surpassing conventional methods.	The CNN model and genetic algorithm depend on training data quality. Computational complexity requires substantial resources. Hyperparameter tuning impacts the genetic algorithm's performance. Limited information hinders assessing potential limitations.
[26, 2019]	Convolution neural network and SVM	Precision 99.02%, recall 98.79%, F-Measurement 98.9%, and MAE (0.78%)	High accuracy rates	Limited ability to generalise
[27, 2018]	Deep learning method	Average gain between 6% and 10% when compared to the method Fast R-CNN	Deep learning algorithms have been shown to achieve state-of-the-art performance on various problems, including image and speech recognition, natural language processing, and computer vision.	This can make it difficult to understand how the model makes predictions and identify any errors or biases.
[28, 2018]	Deep convolutional neural network	It saves 99.3% of the manual labour.	Deep learning models can generalise well to new situations or contexts, as they can learn abstract and hierarchical representations of the data.	Deep learning models can only make predictions based on the data it has been trained on
[29, 2017]	Deep Convolutional Neural Networks for Automated Wildlife Monitoring	It achieved an accuracy of 96.6% for detecting images.	Deep learning models can be easily scaled to handle increasing data and can be deployed on cloud platforms and edge devices.	This can be costly and time-consuming.
[30, 2021]	R-CNN	Accuracy of 90.4 % for detecting most common species.	Autonomous vehicles use it to perceive objects in their surroundings to ensure a safe driving experience.	This sometimes could result in the generation of lousy regional proposals.
[31, 2018]	deep-learning-based method	The error rate is less than 3%	It can see what human minds cannot visualise.	One cannot accurately define the sorting and output of an unsupervised task.
[32, 2023]	re-ID models	Accuracy of 96.6% for *PolarBearVidID* dataset	Developing re-ID models will significantly facilitate the work of biologists and animal caretakers in the future.	Limited dataset and models used
[33, 2020]	ssd_inception_v2 , ssd_mobilenet_v2 , ssd_mobilenet_v2_quantized , ssdlite_mobilenet_v2 and Raspberry Pi	ssd_mobilenet_v2 average precision is high and generates the least number of false positives	MobileNet V2 was selected as the final model for the application. This is due to its traits of generating a small number of false positives and not splitting an event into smaller ones.	MobileNet V2 sacrifices accuracy for speed, potentially leading to missed detections or misclassifications, especially in complex scenes.
[34, 2019]	CNN1 and CNN2	They achieved an accuracy of 95.86% and 98.05% for binary classification. For multiclassification, they achieved accuracies of 83.07% and 90.32%.	The trained CNNs were directly applied to video footage because film can be considered a sequence of image frames. To speed up the detection process, all images were converted to grayscale.	Limited models are used to compare the results.

References for the research on animal detection and recognition using deep learning are based on the following criteria: Relevance to the research topic: The selected papers are directly related to animal detection in videos using deep learning algorithms, which align with the focus of the proposed research. The references cover a range of publication years, indicating the proposed interest in exploring recent advancements and foundational works in the field.

Animal detection models proposed to date have positive and negative aspects [18]. Most models only produce output and predict the results, neglecting to deal with the problem of output unpredictability in terms of accuracy, sensitivity, and specificity [34]. The papers in [1] and [9] have achieved high animal detection accuracy.

- The study in [1] uses a Convolutional Neural Network (CNN) algorithm to detect wild animals, achieving an accuracy of 93.8%. It also mentions automating animal identification with 99.3% accuracy.
- The study in [9] employs Convolutional Neural Networks (CNNs) and achieves an impressive 99.70% training accuracy and 96.85% validation accuracy for real-time monitoring of activities.

Considering the high accuracy and real-time monitoring capabilities, [9] is a promising paper for animal detection using video data.

Data pre-processing is crucial for optimal results. Techniques include noise removal and image enhancement through methods like Butterworth filters, contrast enhancement, and segmentation. Segmentation involves precisely delineating individual animals or their parts within images or video frames. Methods like MSER, Mask R-CNN, and Falzenszwalb algorithm are employed for this purpose. [24] Various models such as Faster-RCNN, Dilated Residual Networks, YOLOv5, and others are used for image classification [26]. Metrics like precision, recall, F1 Score, mean average precision (mAP), intersection over union (IoU), false positive rate (FPR), false negative rate (FNR), and others are used to evaluate the performance of animal detection systems in video data.

Various detection methods like Convolutional Neural Networks (CNN) and others have been employed with different findings and advantages. CNNs have been used for accurate and automated animal identification. The table comparing different models showcases their respective detection methods, primary results, benefits, and disadvantages. Each entry discusses the detection approach, performance metrics, and the strengths and weaknesses of the technique. Some models like Faster R-CNN, Mask R-CNN, Fast R-CNN, YOLOv4 and DeepSORT, use deep learning and tracking for robust animal detection.

These studies demonstrate advanced deep-learning techniques in animal detection, with various models achieving high accuracy rates. However, challenges such as data quality, computational requirements, and model interpretability remain essential considerations. The choice of model depends on the application's specific needs, such as real-time detection, tracking, or species identification. The use of CNNs and its accuracy make it a strong candidate for further research.

V. Limitations and Challenges

A primary challenge was implementing animal identification models in the existing wildlife ecosystem. Although some researchers tested their models on forests, it doesn’t remain easy. A large dataset incorporating many animals would be helpful in training ML and DL models for application in real time. Additionally, due to the computation cost of a vision-based system, it is unlikely to run real-time identification. Advancements can be made by providing a warning in the manner of a message to the neighbouring forest office when the animal is discovered. It can also be used to decrease human-wildlife conflict and animal accidents. Capturing distinctions between creatures from the same family but of different species might be difficult and should be addressed shortly.

Although the approaches utilising deep learning performed well, there are still certain limitations when dealing with submerged multimedia information, such as poor resolution, lighting changes, and intricate backgrounds. More effective approaches for dealing with these issues should be developed based on these limits. Handling more detailed pictures is essential for better understanding animals and their specific activities. Camera and drone technology developments will enable wildlife surveillance at much greater flight heights, reducing interruption to animals in their natural habitats.

While deep learning approaches have demonstrated efficacy, certain limitations persist, especially when handling submerged multimedia data. Factors such as low resolution, fluctuating lighting conditions, and intricate backgrounds pose obstacles that necessitate the development of more robust techniques. Addressing these constraints would yield improved performance in dealing with challenging scenarios. Enhancing the processing of detailed images is pivotal for a more nuanced understanding of animal behaviour and activities. Anticipated advancements in camera and drone technologies hold the potential to facilitate wildlife surveillance at higher altitudes, minimising disturbances to animals in their natural habitats.

VI. Future Challenges and Research Directions

While significant strides have been made in applying deep learning for animal detection on video data, several challenges and avenues for future research emerge. One pressing concern is the need to address the intricacies of adapting models to diverse and dynamic environmental conditions. This includes developing techniques to handle lighting, weather, and vegetation variations, which are paramount for real-world deployment in natural habitats.

Furthermore, the issue of data scarcity remains a formidable obstacle. Future research should explore innovative data augmentation, synthesis, and generation approaches. By creating diverse and representative training sets, models can be more effectively trained to recognise a broader spectrum of animal species and behaviours.

In addition to the challenges, exploring novel data augmentation and synthesis techniques is imperative to mitigate the need for labelled training data. Developing interpretable and explainable deep learning models for animal detection is another critical avenue, enabling researchers to gain insights into model decisions and bolstering trust in automated detection systems. Further, investigating methods for anomaly detection and outlier identification within video data streams can significantly enhance the robustness and reliability of animal detection systems. Additionally, researching federated learning and edge computing approaches holds promise for decentralised deployment in remote and resource-constrained environments. Addressing these multifaceted challenges will advance the animal detection field and foster sustainable coexistence between humans and the animal kingdom.

These include delving into temporal analysis and behaviour modelling for an in-depth understanding of long-term animal behaviour patterns. Exploring multi-species interactions and fine-grained species identification can provide valuable insights into complex ecological relationships. Enhancing real-time adaptability and edge computing capabilities is vital for dynamic environments while addressing ethical considerations and human-wildlife interactions is imperative to ensure responsible technology deployment. The integration of multi-modal data, the development of user-friendly interfaces, and the fostering of interdisciplinary collaboration are crucial for advancing the field. Moreover, research efforts should also focus on privacy-preserving techniques and long-term monitoring for a holistic approach to animal detection. By tackling these multifaceted challenges, researchers can contribute to a more comprehensive understanding of wildlife behaviour and habitat dynamics, ultimately aiding in practical conservation efforts and the coexistence of humans and the natural world.

Considering the increasing importance of interdisciplinary collaborations, future research efforts should seek to bridge the gap between computer vision experts, ecologists, and conservation biologists. A holistic understanding of animal detection technology's impact on ecological research and wildlife conservation can be achieved by pooling expertise from diverse fields.

VII. Conclusion

The traditional method of animal detection relies on direct observation by humans or the utilisation of specialised tools and techniques. It involves visually scanning an area for animal presence, examining tracks and traces left behind, setting up camera trail cameras for remote monitoring, using acoustic devices to capture animal sounds, and employing trail camera ping and capture methods for direct examination. These traditional approaches have provided valuable insights into animal behaviour and ecology, but they can be labour-intensive, time-consuming, and may have limitations in terms of coverage and scalability. Modern technologies are increasingly integrated with traditional methods to enhance animal detection and monitoring capabilities.

Deep learning approaches provide a beneficial tool for continuous tracking and abundance predictions in animal detection, outperforming laborious, individual efforts in only a tiny percentage of the time. The DL algorithms efficiently identify animals with a high degree of accuracy, and the image of the identified animal is shown for improved accuracy so that it may be utilised for other reasons, such as Detecting wild animals in human habitats and preventing wildlife poaching and human-animal conflict.

Researchers would benefit significantly from detecting animals and extracting their characteristics for their research and detailed study of animal species. As a result, the emerging technology of various Machine Learning and Deep learning algorithms may be applied to animal identification and detection. CNN's technique for Detecting large animals in visuals has proven effective.

References

[1] Banupriya, N., Saranya, S., Swaminathan, R., Harikumar, S., &Palanisamy, S. (2020). Animal detection using a deep learning algorithm. J. Crit. Rev, 7(1), 434-439.

[2] Christin S., Hervet É., Lecomte N. Applications for deep learning in ecology. Methods Ecol. Evol. 2019;10:1632–1644. doi: 10.1111/2041-210X.13256.

[3] Zhao Z.-Q., Zheng P., Xu S.-t., Wu X. Object detection with deep learning: A review. IEEE Trans. Neural Netw. Learn. Syst. 2019;30:3212–3232. doi: 10.1109/TNNLS.2018.2876865.

[4] Feng J., Xiao X. Multiobject Tracking of Wildlife in Videos Using Few-shot Learning. Animals. 2022;12:1223. doi: 10.3390/ani12091223.

[5] Mohammad Sadegh Norouzzadeha, Anh Nguyen, Margaret Kosmalac, Alexandra Swanson, Meredith S. Palmer, Craig Packer, and Jeff Clunea, "Automatically identifying, counting, and describing wild animals in camera-trail camera images with deep learning," PNAS, 2018.

[6] C. Chalmers, p. Fergus, c. Aday curbelo montanez, steven n. Longmore, and serge a. Wich on video analysis for detecting animals using convolutional neural networks and consumer-grade drones. (2021).

[7] Mengyu tan, wentao chao, jo-ku cheng on animal detection and classification from camera trap images using different mainstream object detection architectures. (2022).

[8] Łukasz popek, rafał perz, grzegorz galiński on comparison of different animal detection and recognitionon thermal camera images. (2023).

[9] Hussain, a., ali, s., & kim, h. C. (2022). Activity detection for the wellbeing of dogs using wearable sensors based on deep learning. Ieee access, 10, 53153-53163.

[10] Lei, Y., Dong, P., Guan, Y., Xiang, Y., Xie, M., Mu, J., ...& Ni, Q. (2022). Postural behaviour recognition of captive nocturnal animals based on deep learning: a case study of Bengal, slow loris. Scientific Reports, 12(1), 7738.

[11] Lopez-Vazquez, V., Lopez-Guede, J. M., Marini, S., Fanelli, E., Johnsen, E., &Aguzzi, J. (2020). Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. Sensors, 20(3), 726.

[12] Tanishka Badhe, Janhavi Borde, Vaishnavi Thakur on Study of Deep Learning Algorithms to Identify and Detect Endangered Species of Animals. (2022)

[13] Battu, T., & Lakshmi, D. S. R. (2023). Animal image identification and classification using deep neural network techniques. Measurement: Sensors, 25, 100611.

[14] Rajasekaran Thangarasu, Vishnu Kumar Kaliappan, Raguvaran Surendran, Kandasamy Sellamuthu, Jayasheelan Palanisamy, "Recognition Of Animal Species On Camera Trap Images Using Machine Learning And Deep Learning Models," International Journal Of Scientific & Technology Research, 2019.

[15] S. Aich, S. Chakraborty, J.-S. Sim, D.-J. Jang, and H.-C. Kim, ''The design of an automated system for analysing the activity and emotional patterns of dogs with wearable sensors using machine learning,'' Appl. Sci., vol. 9, no. 22, p. 4938, Nov. 2019.

[16] Alexander Loos, Christian Weigel, Mona Koehler, "Towards Automatic Detection of Animals in Camera-Trap Images," European Signal Processing Conference (EUSIPCO), 2018.

[17] Sofia K. Pillai, Dr M. M. Raghuwanshi, Dr P. Borkar, “SuperResolution Mask-R CNN Based Transfer Deep Learning Approach For Identification Of Birds Species,” International Journal of Advanced Research in Engineering and Technology (IJARET), 2020.

[18] Stefan Schneider, Graham W. Taylor, Stefan C. Kremer, “Deep Learning Object Detection Methods for Ecological Camera Trap Data,” Arxiv, 2018.

[19] Atri Saxena, Deepak Kumar Gupta, Samayveer Singh, “An Animal Detection and Collision Avoidance System Using Deep Learning,” SpringerLink, 2020.

[20] Lee, S., Song, Y., &Kil, S. H. (2021). Feasibility analyses of real-time detection of wildlife using UAV-derived thermal and RGB images. Remote Sensing, 13(11), 2169.

[21] Wijeyakulasuriya, D. A., Eisenhauer, E. W., Shaby, B. A., & Hanks, E. M. (2020). Machine learning for modelling animal movement. PloS one, 15(7), e0235750.

[22] Santhanam, S., Panigrahi, S. S., Kashyap, S. K., & Duriseti, B. K. (2021, November). Animal Detection for Road Safety Using Deep Learning. In 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA) (pp. 1-5). IEEE.

[23] Nguyen, H., Maclagan, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., ... & Phung, D. (2017, October). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (pp. 40-49). IEEE.

[24] Trnovszky, T., Kamencay, P., Orjesek, R., Benco, M., & Sykora, P. (2017). Animal recognition system based on convolutional neural network. Advances in Electrical and Electronic Engineering, 15(3), 517-525.

[25] Chandrakar, R., Raja, R., & Miri, R. (2021). Animal detection is based on deep convolutional neural networks with genetic segmentation. Multimedia Tools and Applications, 1-14.

[26] Manohar, N., Sharath Kumar, Y. H., Kumar, G. H., & Rani, R. (2019). Deep learning approach for classification of animal videos. In Data Analytics and Learning: Proceedings of DAL 2018 (pp. 421-431). Springer Singapore.

[27] Mauro dos Santos de Arruda, Gabriel Spadon, Wesley Nunes Goncalves, & Bruno Brandoli Machado, “Recognition of Endangered Pantanal Animal Species using Deep Learning Methods,” IJCNN, 2018.

[28] Mohammad Sadegh Norouzzadeha, Anh Nguyen, Margaret Kosmalac, Alexandra Swanson, Meredith S. Palmer, Craig Packer, and Jeff Clunea, “Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning,” PNAS, 2018.

[29] Hung Nguyen, Sarah J. Maclagan, Tu Dinh Nguyen, Thin Nguyen, Paul Flemons, Kylie Andrews, Euan G. Ritchie, and Dinh Phung, “Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring,” Deakin University, Geelong, Australia, 2017.

[30] F. Schindler and V. Steinhage, “Identification of animals and recognition of their actions in wildlife videos using deep learning techniques,” Ecological Informatics, vol. 61, p. 101215, Mar. 2021, doi: https://doi.org/10.1016/j.ecoinf.2021.101215.

[31] T. D. Pereira et al., “Fast animal pose estimation using deep neural networks,” Nature Methods, vol. 16, no. 1, pp. 117–125, Dec. 2018, doi: https://doi.org/10.1038/s41592-018-0234-5.

[32] M. Zuerl et al., “PolarBearVidID: A Video-Based Re-Identification Benchmark Dataset for Polar Bears,” Animals, vol. 13, no. 5, p. 801, Jan. 2023, doi: https://doi.org/10.3390/ani13050801.

[33] Amanda Tydén and Sara Olsson, “Edge Machine Learning for Animal Detection, Classification, and Tracking”, 2020.

[34] R. Chen, R. Little, L. Mihaylova, R. Delahay, and R. Cox, “Wildlife surveillance using deep learning methods,” Ecology and Evolution, vol. 9, no. 17, pp. 9453–9466, Aug. 2019, DOI https://doi.org/10.1002/ece3.5410.

[35] Jackulin Mahariba, A. U. (2022). An efficient automatic accident detection system using inertial measurement through machine learning techniques for powered two wheelers. Expert Systems With Applications, 0957.

[36] Ahmed Yaseer, H. C. (2021). A Review of Sensors and Machine Learning in Animal Farming. 11th IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (p. 21). Jiaxing, China: EEE Xplore.

[37] Dario Augusto Borges Oliveira, L. G. (6 September 2021). A review of deep learning algorithms for computer vision systems in livestock. Livestock Science,1-15.

[38] Heegon Kim, J. S. (2015). Automatic Identification of a Coughing Animal using Audio and Video Data. ISCC 2015 (p. http://pos.sissa.it/). Guangzhou, China: ISCC.

[39] Jun Bao, Q. X. (2022). Artificial intelligence in animal farming: A systematic literature review. Journal of Cleaner Production, 0959.

[40] Md Ekramul Hossain, M. A. (2022). A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. Artificial Intelligence in Agriculture, 40.

[41] Md Sultan Mahmuda, A. Z. (2021). A systematic literature review on deep learning applications for precision cattle farming. Computers and Electronics in Agriculture, 0168.

[42] Prashanth C. Ravoor, S. T. (2020). Deep Learning Methods for Multi-Species Animal Re-identification and Tracking – a Survey. Computer Science Review,0137.

[43] Qiumei Yang, D. X. (2020). A review of video-based pig behavior recognition. Applied Animal Behaviour Science,1-7.

[44] Rodrigo García, J. A. (2020). A systematic literature review on the use of machine learning in precision livestock farming. Computers and Electronics in Agriculture,0168.

[45] S Jeevitha, D. V. (May 2020). A Review of Animal Intrusion Detection System. International Journal of Engineering Research & Technology,129-1221.

[46] Shi Dong, P. W. (2021). A survey on deep learning and its applications. Computer Science Review, https://doi.org/10.1016/j.cosrev.2021.100379.

[47] Verma, A. D. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, Progress in Artificial Intelligence.

[48] Vigneshwaran Palanisamy, N. R. (2021). Detection of wildlife animals using deep learning approaches: A Systematic review. 21st International Conference on Advances in ICT for Emerging Regions (ICTer 2021) (pp. 153-158). India: IEEE Explore.

[49] Vipal Kumar Sharma, R. N. (11 September 2020). A comprehensive and systematic look up into deep learning based object detection techniques: A review. Computer Science Review,1574-0137.

[50] Weinstein, B. G. (2017). A computer vision for animal ecology. Journal of Animals Ecology, 533-545.

[51] A Literature Research Review on Animal Intrusion Detection and Repellent Systems. (2021). L. Ashok Kumar, R. Neelaveni, M. Kathiresh, P. Sweety Jose, N. Archana, S. Saravanakumar, 227.

[52] Ferrante, G. S., Rodrigues, F. M., Andrade, F. R., Goularte, R., & Meneguette, R. I. (2021). Understanding the state of the Art in Animal detection and classification using computer vision technologies. 2021 IEEE International Conference on Big Data (Big Data) (p. https://doi.org/10.1109/BigData52589.2021.9672049). Orlando, FL, USA: IEEE Xplore.

[53] M.Sowmya, D. M. (2021). A Review On Animal Detection Using Different Detection Techniques. Turkish Online Journal of Qualitative Inquiry (TOJQI), 8249 - 8254.

[54] Sreedevi C K, S. E. (2019). Automated Wildlife Monitoring Using Deep Learning. In Proceedings of the International Conference on Systems, Energy & Environment (p. 7). Kerala: SSRN.

[55] L. Ashok Kumar, R. N. (2021). A Literature Research Review on Animal Intrusion Detection and Repellent Systems. ICCAP 2021, (p. 227). Chennai, India.

[56] Manasa Kommineni, M. L. (2022). Agricultural Farms Utilizing Computer Vision (AI) And Machine Learning Techniques For Animal Detection And Alarm Systems. JOURNAL OF PHARMACEUTICAL NEGATIVE RESULTS, https://doi.org/10.47750/pnr.2022.13.S09.411.