A Comprehensive Survey of Animal Identification: Exploring Data Sources, AI Advances, Classification Obstacles and the Role of Taxonomy

Wiley
International Journal of Intelligent Systems
Volume 2024, Article ID 7033535, 25 pages
https://doi.org/10.1155/2024/7033535

WILEY

Review Article

A Comprehensive Survey of Animal Identification: Exploring Data Sources, AI Advances, Classification Obstacles and the Role of Taxonomy

Qianqian Zhang, Khandakar Ahmed, Nalin Sharda, and Hua Wang

Institute for Sustainable Industries and Liveable Cities (ISILC), Victoria University, Footscray 3011, Victoria, Australia

Correspondence should be addressed to Qianqian Zhang; yindie36@gmail.com and Khandakar Ahmed; khandakar.ahmed@vu.edu.au

Received 21 March 2024; Revised 1 June 2024; Accepted 11 September 2024

Academic Editor: Zhiyuan Qi

With the rapid development of entity recognition technology, animal recognition has gradually become essential in modern society, supporting labour-intensive agriculture and animal husbandry tasks. Severe problems such as maintaining biodiversity can also benefit from animal identification technology. However, certain invasive recognition systems have resulted in permanent harm to animals, while noninvasive identification methods also exhibit certain drawbacks. This paper conducts a systematic literature review (SLR), presenting a comprehensive overview of various animal recognition technologies and their applications. Specifically, it examines methodologies such as deep learning, image processing and acoustic analysis used for different animal characteristics and identification purposes. The contribution of machine learning to animal feature extraction is highlighted, emphasising its significance for animal taxonomy and wild species monitoring. Additionally, this review addresses the challenges and limitations of current technologies, including data scarcity, model accuracy and computational requirements, and suggests opportunities for future research to overcome these obstacles.

Keywords: animal identification; image processing; machine learning; neural network; signal processing

1. Introduction

The globe is facing the sixth mass extinction. Understanding and protecting biodiversity is one of the biggest challenges for humankind in the face of animal habitat loss, environmental pollution, climate change and other factors that have led to a sharp decline in biodiversity in the past few decades [1]. Consequently, half of the species may become extinct in the next 100–300 years [2]. Moreover, many species, especially those less viable and challenging to monitor, were extinct before they were correctly described. Some of these have economic significance to human society or are an essential part of the natural environment, and their extinction is a tragic permanent loss [3]. Interestingly, according to [1], nonprofessional amateurs describe more than 60% of the unknown species, i.e., the complete spectrum of known species has not all been counted by biologists systematically. The unprofessional labelling has also led to the incorrect recording of some species and, at times, multiple ways of describing the same species. Hence, bio-taxonomy is often described as science in crisis due to the lack of workforce and funding. A study has shown that if a camera trap generates 700,000 images annually, staff may spend nine months classifying the images based on existing species [4]. In addition, according to the real-time data of the open-source biodiversity data website, more than 1.3 billion pictures of the phylum Chordata have not been classified [5], let alone those insects prone to mutation due to their physiological structure. Moreover, the traditional standard taxonomy requires complex laboratory equipment, and even professional biologists must spend much time performing classification. Systematic and efficient resource management and identification systems are necessary to alleviate this limitation. This article conducts a systematic literature review (SLR) on animal recognition. The SLR selected 54 core publications of interest from January 2017 to December 2023 according to the Kitchenham methodology [6].

2. Background

This section introduces some background concepts on which the SLR is based. First, Sections 2.1 and 2.2 present general animal identification applications from two perspectives: invasive and noninvasive. Next, the application of individual identification and the current challenges of species identification are introduced in Section 2.3. Finally, Section 2.4 analyses the existing literature reviews and surveys on animal identification.

2.1. Invasive Methods. The earliest record of animal body marking and identification is the Code of Hammurabi, 3800 years ago [7]. Ever since humans have kept animals, ownership marks have been used on animals as a fundamental need to this day. After thousands of years of development, people in modern society identify and track animals as individuals or in groups using different marking methods for commercial or research purposes, including commodity traceability, bio-security and species tracking. Some standard invasive techniques are as follows:

- Implantable devices: microchip and acoustic tag
- Nonimplantable devices: ear tag and wing tag
- Permanent damage: earmarking, toe clipping, fin clipping and manual tattoos (lip, tail, footpad or ears)

Most ancient civilisations applied traceability records of permanent body markings to valuable animals and kept written records [8]. For example, Alexander the Great and the Chinese in the seventh century used red-hot irons to mark the chest or group of horses and oxen and combined them with name tags as official registration information to determine the ownership of the animals [7]. This way of destroying skin tissue poses a potential risk of infection and is difficult to implement in long-haired animals. Also, the branding of certain parts is not permanent and needs to be reimplemented occasionally on body parts such as horse hooves and sheep horns. Compared to traditional hot-iron branding, freeze branding is almost painless. It destroys the pigment-producing hair cells, causing the hair to grow out of the mark to be white. This method can also be applied to hairless animals such as aquatic life [9].

Another widely used method is tattoos and clipping. Like branding, the tattoo is applied without anaesthesia by adding ink or dye into the wound to present specific numbers, words or patterns. Clipping means removing a body part of an animal to create a particular shape or piercing to wear a tag. It usually eliminates components that do not affect functionality, such as the edge of the ear, fish fins, bird wings tips and lab mice toes [10]. While tattoos are generally permanent on domestic animals, they may fade in wild animals due to cell shedding with age. Clipped parts do not regenerate, making them a more reliable method for identification purposes.

The limitations of traditional identification methods like tattooing and clipping, which inflict bodily harm and carry limited data, have led to the adoption of radio-frequency identification (RFID) technology. RFID, comprising a small radio transponder, receiver and transmitter, gained popularity in the 1970s. Three primary implantable RFID tags for animal identification are boluses (retained in ruminant stomachs), ear tags and injectable glass tags [11]. RFID's large data capacity allows for the creation of unique animal profiles, facilitating the tracking of individual movements, lineage and ownership information. This technology enables the comprehensive management of farm animal data, including production metrics, growth rates, health records and breeding details [12]. In wildlife research, RFID tags assist in studying survival patterns and migrations [13]. Some countries also mandate microchipping pets for easy identification if lost [14]. While RFID offers easier readability and management and reduces animal stress, it is costlier than traditional methods and poses potential food safety risks due to residual tags [15].

With the demand for laboratory animals and the recognition of animal welfare, the Three Rs (3Rs) principle is receiving more attention. It was first proposed by Russell and Burch [16] and provided a framework for ethical decision making when using animals in research and teaching. 3Rs specifically refer to

- Replacement: Where possible, consider replacing or partially replacing animals, or use nonsentient material, such as higher plants and microorganisms, to replace conscious living vertebrates.
- Reduction: Minimise the number of experimental animals and use the same animal in multiple experiments.
- Refinement: Avoid or minimise pain, fear and lasting harm.

Based on the above three principles as a framework, each country has its own rules to optimise research on animals' health and well-being. On the other hand, it is also considered essential to avoid unreliable scientific results such as low-quality data and confusing behaviour [17]. Some data show that under the influence of emotions such as fear and pain, some animals will cause deviations in experimental results (such as movement trajectories, vocalisations, behaviours and physical functions) [17].

Therefore, with the advancement of computer image processing technology, more noninvasive animal monitoring and identification methods that can reduce the interference of individual animals have been proposed.

2.2. Noninvasive Methods. Compared to those invasive methods, noninvasive technologies are more animal-friendly. This section covers wearable equipment, computer vision–based animal identification technologies and other noninvasive techniques.

2.2.1. Wearable Equipment and Dyeing. Foot rings have also worked for birds over the centuries. After the widespread use of RFID, some manufacturers embedded chips in collars to replace the traditional equipment engraved with serial numbers or identity information. These wearable devices containing identifiable information have contributed outstandingly to preventing and controlling rabies and other animal diseases [18]. However, the drawbacks of these wearable devices are also evident. For example, Global Positioning System (GPS) is not adequately accurate for farm animals. Although RFID can store rich individual information, it relies heavily on manual input; therefore, most wearable devices have high installation and maintenance costs [19].

On the other hand, semipermanent marking dye on animal hair will naturally fall off during the natural hair growth process without causing any distress to the animal. The drawbacks of this approach are also obvious—limited information and complex management. Nowadays, noninvasive methods are gradually replacing invasive procedures and are being accepted by modern society, especially with the development of artificial intelligence (AI) technology.

2.2.2. Computer Vision and Image Recognition. Artificial neural network (ANN) imitates the central nervous system of animals, and it is also a nonlinear statistical data modelling tool. ANNs have gradually been used in many applications prevalent in modern society and have brought substantial advancements in computer vision applications. Applying ANNs to biology research has led to the development of many identification methods [20]. These methods reduce the impact on the animal and provide greater flexibility and efficiency. Computer vision–based commonly used methods include the following:

- Static feature recognition (SFR): footprint, pattern, muzzle, iris, ear blood vessels and wing texture.
- Dynamic feature recognition (DFR): actions and migration trajectories.
- Thermal recognition (TR).
- Real-time monitoring (RTM): drone aerial imagery and GPS.

Machine learning advances, particularly deep learning, provide powerful support for accurate image-based recognition. Like face recognition, ANNs support universality, uniqueness, permanence, measurability, feasibility and reliability, which are the basis for selecting identification features [21]. It allows feature extraction for animals with various appearances and body structures.

2.3. Individual Identification and Group Identification. Individual animal identification is essential to applying ANNs for biological research [20]. Some small-scale individual identification projects have already been put into commercial use. For instance, Mathieu et al. proposed a model based on convolutional neural network (CNN) and deep belief network (DBN) [22]. Using this method, farm managers can carry out disease monitoring and food control regimes by extracting the facial features of domestic animals and observing subtle changes in the pig's appearance without any physical interaction with the pig [22].

ANNs have received attention for their ability to distinguish between species. Since the last century, the camera trap technique has received wide acceptance as it provides a large amount of image data for observing biodiversity and monitoring population size. More recently, cost and time have been saved further since drones have been used for field image collection. However, the traditional manual screening is labour-intensive and time-consuming. A study shows that the accuracy of wild animal population estimates from aerial images is highly correlated with funding, labour, time and human skills [23]. Therefore, a tool that can quickly and efficiently classify animals for their images is urgently needed. However, most existing research on species classification is still in the experimental stage. Some of the proposed system architectures are not scalable to large-scale datasets because the experimental objects contain only a small number of classes. Some cannot achieve a satisfactory recognition rate due to the lack of source images.

2.4. Literature Reviews and Surveys on Animal Recognition. Although automated information processing technology has flourished in the past two decades and people have gradually paid more attention to biodiversity, there are few SLRs for animal identification. Most are analyses of specific techniques or have little relevance to animals.

Animal identification is sometimes presented as a small part of a study. The statistical results proposed by Disney et al. realise that improving animal identification can effectively reduce the losses caused by animal diseases [24].

Some studies focus on one species or class of organisms. A series of bioscience techniques used to detect and monitor aquatic animals are listed in the survey by Rees et al. [25]. Jukan et al. presented a literature review focussing on the impact of intelligent computing and sensing technologies on animal welfare [26]. They analysed different recognition techniques rather than recognition rates on domestic, farm and wild animals.

Schneider et al. summarised technologies and applications of animal re-identify. Reidentify refers to recording a specific individual target and accurately identifying when it enters the surveillance range again [27]. It is mainly used in the tracking and monitoring of individual animals. In contrast, object classification is usually used to categorise items that have never been seen before. The use of computer vision to reidentify animals dates back to 1990 [28]. Schneider et al. listed feature-engineered and deep learning approaches for animal reidentification.

On this basis, Ravoor added open set recognition, multispecies reidentification and applicability to automation for tasks [29]. In addition to reviewing publications on object tracking, they also investigated the uses of deep learning for feature extraction. Another survey on wildlife identification summarises data collection methods and computer-assisted techniques for individuals and species [30]. It is worth noting that Petso et al. listed previous studies according to different identification attributes, and each attribute is divided into two categories: species and individuals. However, these surveys have focussed on the technology that empowers AI while ignoring other widely used or historically significant identification methods. In addition, almost all of the current animal recognition survey papers are more inclined to study computer science while ignoring the role of biology in animal recognition.

3. Planning and Conducting the SLR

This section introduces SLR planning according to Kitchenham's theory [6]. Each step of the SLR is described in detail to improve reproducibility. This paper discusses publications on animal identification research in Section 4.

3.1. Planning the SLR. The SLR planning includes the following steps:

1. Research scope presentation and articulation of the survey questions.
2. The search string identification.
3. Data source selection.
4. Inclusion criteria definition.

3.1.1. Survey Questions. The main goal of the SLR is to present an overview of the animal identification technology in practice and to articulate how to choose the most suitable animal feature extraction method for different identification needs. With this goal in mind, the following survey questions are formulated:

1. SQ1: What data sources and methods have researchers investigated so far in the literature for animal identification studies?
2. SQ2: How has AI, with the recent development of deep learning, changed the dynamics of the identification and recognition landscape, particularly the impact on animal identification research?
3. SQ3: How do scientists process large-scale datasets for animal classification?
4. SQ4: What are the limitations of animal classification techniques when applied to a wide range of applications?
5. SQ5: What is the significance of taxonomic rank in animal identification?

SQ1 provides an overview of popular animal identification methods and classifies them into four categories based on different data sources: audio, video, image and wireless signals. The success of these studies has to be attributed to machine learning. SQ2 discusses aspects of machine learning that make contributions to animal identification applications. The success of these machine learners is also dependent on suitable datasets. In the literature, some of them are handy in large-scale datasets. SQ3 introduces the research into multiple animal species with large-scale datasets. However, animal classification is still a novel topic. Due to various reasons, many animal classifications face some problems and have not been widely spread. SQ4 discusses current or future challenges and shortcomings of animal classification. And back to the core of this survey, a new perspective for animal identification or classification systems is provided. SQ5 summarises the relationship between biological taxonomic rank and animal identification, focussing on analysing the importance of taxonomic rank in animal identification.

3.1.2. Definition of the Search Strings. Twenty-four search strings are defined by deriving terms from (1) an initial assessment of some of the most cited papers dealing with animal recognition technology and (2) subject matter knowledge. First, the term 'animal' is included in each set of strings to focus on animal-related publications and block articles related to humans or other objects. After that, the differences in recognition, identification and classification terms are compared:

- Animal recognition: The animal behaviour, one or more characteristics match the data in the database.
- Animal identification: The identification information of the individual animal was identified.
- Animal classification: Extract animal characteristics and classify animals according to bio-taxonomy.

To limit the scope of the search in the fields of engineering and computer science, the above-listed terms are combined with terms such as 'intrusive', 'RFID', 'object detection', 'machine learning', 'deep learning', 'convolutional neural network' and 'taxonomy rank'. It is worth noting that 'CNN', 'convolutional neural network' and 'convolutional neural networks' are used simultaneously, as an important string because abbreviations and their longer versions may return different results.

According to the criteria given above, the final search strings were the following:

1. Intrusive and '*x*'
2. 'Object detection' and '*x*'
3. 'Machine learning' and '*x*'
4. 'Deep learning' and '*x*'
5. 'CNN' and '*x*'
6. 'Convolutional neural network' and '*x*'
7. 'Convolutional neural networks' and '*x*'
8. 'Taxonomic rank' and '*x*'

'*x*' indicates 'animal recognition', 'animal identification' and 'animal classification' for a total of 24 strings. For example, the first three strings are 'intrusive and animal recognition', 'intrusive and animal identification' and 'intrusive and animal classification'.

3.1.3. Selection of Data Sources. Two data types are considered for this SLR: scientific digital libraries and selected conferences and journals. The selected scientific digital libraries are ACM Digital Library, Scopus, IEEE Xplore Digital Library, Web of Science Core Collection and Springer Link. All searches are performed on the entire database based on the search strings. Since the research covers multiple disciplines, publications related to animal recognition span various fields, such as computer science, engineering, agricultural and biological sciences.

Ten specific conferences and four important journals are also selected as complementary data sources for the SLR due to their high reputation and vital relevance to the scope of the paper:

3.1.3.1. Conferences

1. International Conference on Communication and Electronics Systems (ICCES)
2. International Conference on Information and Computer Technologies (ICICT)
3. International Conference on Electrical, Computer and Communication Technologies (ICECCT)
4. International Conference on Computer Communication and the Internet (ICCCI)
5. International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)
6. International Conference on Information, Intelligence Systems and Applications (IISA)
7. International Conference on Communications (ICC)
8. Symposium Series on Computational Intelligence (SCSE)
9. International Conference on Computer Science and Software Engineering (CSASE)
10. Conference on Information and Communication Technology (CICT)

3.1.3.2. Journals

1. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
2. Pattern Recognition
3. IEEE Transactions on Image Processing
4. Nature Machine Intelligence and Sensors

3.1.4. Definition of Inclusion Criteria. Finally, Table 1 shows the inclusion criteria used to refine the results obtained from the search engines to ensure an unbiased selection of relevant publications.

3.2. Conducting the SLR. As suggested by Kitchenham, literature review execution and data synthesis are conducted after planning and are presented in Sections 3.2.1 and 3.2.2.

3.2.1. Literature Review Execution. This section performs the search process, which contains three main phases:

1. Phase 1—digital library search: Search each digital library indicated in Section 3.1.3 by the search string described in Section 3.1.2.
2. Phase 2—conference and journal search: Search journals and conferences mentioned in Section 3.1.3 using the search strings described in Section 3.1.2.
3. Phase 3—backward snowballing search: Examine the bibliography of publications produced in the first two phases to identify possible relevant publications according to the method proposed by Wohlin [31].

A total of 3096 potentially relevant publications emerged. The number of publications produced by searching each digital library is available on https://cutt.ly/B0sly5z.

After removing duplicates, a dataset containing 1447 publications is obtained. During a rough count, the nearest inflexion point in the number of articles on animal identification was identified in 2019. To avoid an explosion of similar types of articles due to a breakthrough technology at a particular time, the search year is expanded to the second inflexion point, 2017. Therefore, the pool is reduced to 969 manuscripts. These 969 publications were reviewed individually for their titles, abstracts, introductions and conclusions according to the inclusion criteria mentioned in Section 3.1.4. Sometimes, the full paper is reviewed to determine its appropriateness for the SLR. Finally, the first phase resulted in a collection of 33 publications.

In Phase 2, the conference proceedings and journals specified in Section 3.1.3 are searched using the search strings given in Section 3.1.2. The timeline of this search is limited to 2017 onwards. The search results yielded 47 potentially relevant publications. The specific number of search results is accessible at https://cutt.ly/r0szTiA. These manuscripts are reviewed, and inclusion criteria are applied, resulting in five publications of relevance. The second set is merged with the set obtained in the Phase 1 search. After removing two duplicates, the total tally turns to 36.

Phase 3 uses the backward snowballing method [31], which identifies new papers by reviewing the references and citations in the publications already obtained. The snowballing phase yielded 18 additional publications—the final 54 articles are listed in alphabetical order of the first author's last name in Table 2. Detailed information is available as an Excel sheet at https://cutt.ly/5wN7XXgt. Figure 1 shows the process followed in the three search phases.

3.2.2. Data Synthesis. The results of the abovementioned search process present the most critical contemporary knowledge. The following three factors ascertain the veracity of the search. First, the three phrases—animal identification, classification and recognition—return different results in some cases. All three must be used for retrieval to ensure no valuable research work is missed. Second, three terms are used for a 'convolutional neural network': the acronym 'CNN', 'convolutional neural network' and 'convolutional neural networks' in the plural. These terms return slightly

Table 1: Inclusion criteria list.

(1) The publication deals with technology-related issues about animal recognition (publications that mention animal identification only for economic benefit analysis are excluded)
(2) The publication is a high-rank journal based on SCImago Journal Rank (SJR)
(3) The publication has been published after 2016
(4) The publication has been peer-reviewed
(5) The publication is written in English

different results. As the core of this topic, all three are used to ensure no publications are missed. Third, Figure 2 presents that there has been a stable trend in the level of attention within the field of animal recognition from 2018 to 2021, closely aligning with the years searched.

4. Reporting and Analysing the SLR Results

This section presents the analysis of the SLR results considering the four survey questions articulated in Section 3.1.1. Each subsection addresses multiple aspects aiming to analyse the problem from different perspectives.

4.1. Methods to Identify Animals Based on Different Data Sources. Since humans started keeping animals in captivity, animal identification has been essential. Section 2 describes the animal identification approaches in different eras. However, with the development of technology, it is gradually replaced by more efficient and convenient computer vision. As an essential part, the data source is a medium that stores related data in different formats. Generally speaking, different sources of data store various characteristics, and researchers will choose a suitable type of source for different research purposes. For example, the image data store information about the animal's appearance, the voiceprint information stores the different calls of the animal and the thermal energy information stores the movement track information of the individual animal in the process of migrating. This section analyses 54 articles retrieved in Section 3.2 based on the following four categories of data sources:

- Audio
- Image
- Video
- Wireless signal

These publications are listed alphabetically by author's last name in Table 2. Table 3 presents the same list categorising these publications based on the identification method used for retrieving these. Further details are presented in the following sections.

4.1.1. Audio. Audio signals can be represented in either analogue or digital format. Early audio signal processing included the transmission and storage of inventions such as the landline telephone, the phonograph and the radio. Until the widespread use of digital technology, digital audio signal processing was the first choice because it was more powerful and efficient.

Thangavel and Shokkalingam provided an animal recognition study based on audio signal processing [70]. The main contribution of their research is that they used the Mel Frequency Cepstral Coefficient (MFCC) to extract animal voiceprint features from audio signals to identify species and used the Gammatone Frequency Cepstral Coefficient (GFCC) to detect the psychological state of elephants. Also, Thangavel et al. demonstrated other roles of acoustic signals in animal recognition, such as detecting the age, gender, social group or kinship group of animals. However, the cognition rate in these aspects is not as satisfactory as species recognition, and the correct rate is about 75%. In addition, compared to visual signals, the information collected by sound signals can be far away from the detector and is less affected by the environment. In this study, the sounds of a herd of elephants were recognised from 50 m or even several kilometres away and alerted by devices linked to an Internet of Things (IoT) system.

Similarly, another study presents that sounds from animal movements were collected to detect specific behaviour. Nasirahmadi et al. [35] collected 9200 turkey pecking sounds for training to detect turkey quality and prevent disease. The researchers noted that one of the challenges of using CNNs for acoustic recognition is visualising the audio signal. Although an overall accuracy of 93.3% was achieved, this project needs to investigate the effect of other factors, such as birds' age and breeding conditions.

While insects represent most animal species, data are particularly lacking [82]. This is due to their living environment and behavioural habits that make it challenging to capture their pictures. The significance of collecting invertebrates, especially insects, for monitoring and classification is summarised in a study by Høye et al. [69]. First, the researchers collected image data for insects with phototaxis by placing artificial light and arranging traps to attract insects. Second, for insects that are difficult to film, acoustic signals, such as the sound of flapping wings, can be collected to classify species. The researchers point out that deep learning is a valuable tool for classifying insect acoustic information.

Acoustic monitoring is more complicated than intuitive camera-captured images because it usually takes a lot of time to sample. Therefore, it is important to train the model with existing datasets. The Orchive is a 19,000-hour dataset that was recorded over a period of 23 years of killer whale sounds [83]. Bergler et al. used over 10,000 labels from the Orchive

Table 2: List of the 54 publications selected by the SLR from January 2017 to December 2023.

Authors	Reference
Arruda et al. (2018)	[32]
Bergler et al. (2019)	[34]
Cao et al. (2019)	[36]
Chen et al. (2019)	[38]
Daniel et al. (2019)	[40]
Delplanque et al. (2021)	[42]
Demir et al. (2020)	[44]
Ditria et al. (2020)	[46]
Divya et al. (2019)	[48]
Eikelboom et al. (2019)	[23]
El et al. (2020)	[51]
Fang et al. (2019)	[53]
Ferreira et al. (2020)	[55]
Gray et al. (2019)	[57]
Guo et al. (2019)	[59]
Guzman et al. (2021)	[61]
Han et al. (2019)	[63]
Han et al. (2020)	[65]
Hong et al. (2019)	[67]
Høye et al. (2021)	[69]
Kellenberger et al. (2019)	[71]
Kellenberger et al. (2018)	[73]
Kumar et al. (2018)	[75]
Lopez et al. (2023)	[76]
Miele et al. (2021)	[78]
Marsot et al. (2020)	[22]
Moallem et al. (2021)	[4]
Moreira et al. (2017)	[33]
Nasirahmadi et al. (2020)	[35]
Norouzzadeh et al. (2018)	[37]
Nguyen et al. (2017)	[39]
Okura et al. (2019)	[41]
Psota et al. (2019)	[43]
Ramesh et al. (2019)	[45]
Rauf et al. (2019)	[47]
Rodriguez et al. (2018)	[49]
Ruff et al. (2021)	[50]
Saleh et al. (2018)	[52]
Schneider et al. (2020)	[54]
Schwartz et al. (2021)	[56]
Shahinfar et al. (2020)	[58]
Shepley et al. (2021)	[60]
Singh et al. (2021)	[62]
Tabak et al. (2019)	[64]
Tabak et al. (2018)	[66]
Tan et al. (2020)	[68]
Thangavel et al. (2021)	[70]
Ulhaq et al. (2021)	[72]
Van et al. (2018)	[74]
Vidal et al. (2021)	[20]
Willi et al. (2019)	[77]
Xu et al. (2020)	[79]
Yousif et al. (2018)	[80]
Zhao et al. (2021)	[81]

Annotation Catalog (OAC) corpus and over 30,000 whale samples from DeepAL Fieldwork Data 2017 and 2018 (DLFD) to train, validate and test the model [34]. They used PyTorch to implement data preprocessing training on the ORCA-SPOT architecture and explored the feasibility of deep learning methods to study animal communication based on the songs of humpback whales (*Megaptera novaeangliae*). However, acoustic classification still faces many challenges. Another study presents that water acoustic signals are influenced by factors such as turbidity, salinity and temperature [84]. These factors allow the acoustic signal to propagate along a curve in water rather than in a straight line as it propagates in air. On the other hand, some same species can even have ‘dialects’ due to regional differences [85].

The other study on acoustic classification is not only the identification of single species but also the identification of multiple species in the same environment. Ruff et al. developed a deep neural network (DNN) to recognise sounds made by 14 animals [50]. They split the audio clips into audio spectrograms, each representing 12 s. Unlike image augmentation for image recognition, spectrogram augmentation generally applies the randomised offset and dynamic range. They compiled and trained a CNN model with four convolutional layers and two fully connected layers. They then tested the model with an independent test set containing noise, one or more species. Even if the recognition rate of the final model in the natural environment does not reach the ideal efficiency, high-performance computer recognition still saves a lot of time and labour costs.

Furthermore, Xu et al. proposed a framework in which local processing is performed by the Wireless Acoustic Sensor Network (WASN), and DNNs are run on the server side to identify species [79]. They used two datasets, one containing 14 species of frogs and the other containing 20 crickets. The researchers obtained segments, including only animal acoustic signals, by removing not-of-interest areas from the background. Their contribution is that this framework can solve the cost of local hardware and speed up the recognition rate to achieve the purpose of RTM. Table 4 summarises related information of the six acoustic identification publications from the 54 target papers.

4.1.2. Image. Many animal recognition systems use images as source data. As mentioned in Section 2.2.2, computer vision emerged as a powerful technology towards the end of the last century. In particular, the high-performance graphics processing unit (GPU) has played an important role in image recognition since it was invented. This section will analyse the applications of image identification based on different types of images. Depending on the extracted features, the images are categorised into animal appearance, behaviour analysis, group migration and thermal imaging. In general, different animal features are extracted for different research purposes. Figure 3 presents the relationship between various research aims, the four categories and the effects of the environment.

4.1.2.1. Animal Appearance. The features of an animal’s appearance are the most common method to recognise it. Some unique visual characteristics are often the first factors used in recognising them. Kumar and Singh proposed

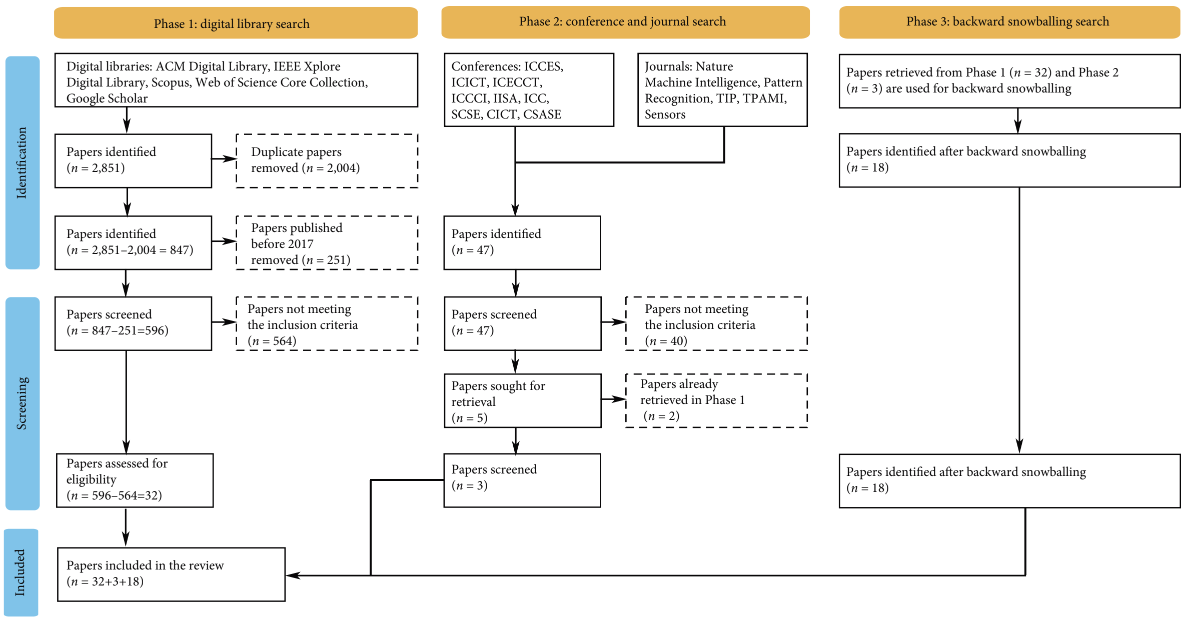


Figure 1: Flow diagram summarises the publication selection along the three search phases.

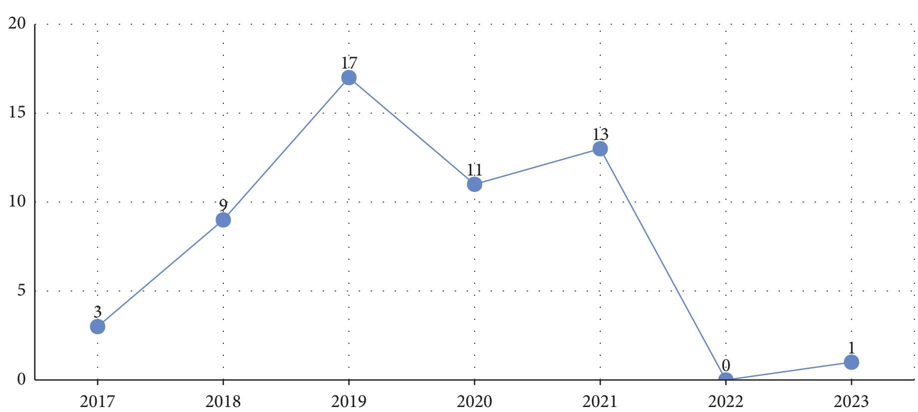


Figure 2: Distribution over time of the publications selected by the SLR.

a Fisher Linear Projection and Preservation (FLPP) feature extraction and representation method to identify and monitor pet dogs in smart cities [75]. However, the recognition rate was not satisfied because it was affected by ambient light. On the other hand, after the texture features are extracted using Local Binary Pattern (LBP) and Speed Up Robust Transform (SURF), they can be encoded into binary patterns to get a better match for the face image. Experimental results show that it takes 10.25 s for the system to recognise a single dog in a database of 5000 images, which is not satisfactory for many real-time applications. There will be delay when using larger datasets.

Ditria et al. proposed a fish detection system based on the mask region–based convolutional neural network (R-CNN) framework to help ecologists [46]. The researchers split video clips taken by submerged action cameras, collated the images containing the target fish (luderick) into a dataset and labelled them. They quantified the models according to different accuracy and recall rates and trained three models with random subsets. It is worth noting that this study compared human beings with deep learning models. The researchers invited two groups of people as control groups. One includes undergraduate students in related disciplines and people interested in fish. And the other one consisted of fish scientists with PhDs. Undoubtedly, computer models performed far faster than both human groups but had the highest rate of false negatives. Human beings can then improve recognition accuracy by framing before and after when the image is blurred, which computer models lack.

Norouzzadeh et al. demonstrated a deep learning architecture to automate multiple tasks on a large dataset [37]. These tasks include identifying animal species, trying to count animals in pictures, describing their behaviour and identifying whether there are cubs. Comparing nine popular DNN architectures, they achieved around 93% accuracy on the world's largest wildlife dataset—Snapshot Serengeti (SS) dataset [86]. Figure 4 presents a part of the samples of the SS dataset.

Table 3: Categories of different animal identification methods based on different source data.

Categories	Included papers	References
Audio	Papers that focus on animal sound analysis techniques	[34, 35, 50, 69, 70, 79]
Image	Articles that focus on computer image processing to improve the identification rate	[4, 20, 22, 23, 32–34, 36–40], [42–69, 71–81]
Video	Papers that focus on audio-visual data sources to identify animals	[38, 40, 41, 46, 72, 81]
Wireless signal	Papers that use wireless signal processing to detect animals	[36, 70, 72]

Table 4: A summary of six acoustic identification publications.

Ref.	Sound category	Identification types	Animal species	Research aims	Methods
[34]	Animal's song	Individual identification	Humpback whale	The meaning of their songs	ORCA-SPOT network architecture
[35]	Pecking activity	Individual identification	Turkey	Disease control	Log-mel spectrogram
[50]	Animal's song	Species classification	14 animal species	Species classification	Spectrograms
[69]	Wing vibration	Species classification	2 species	Collect insect information	Vertical-looking radars (VLRs)
[70]	Pecking activity	Individual identification	Elephant	Age, gender, social group, relationship and behaviour analysis	MFCC, GFCC, SVM
[79]	Animal's song	Species classification	14 frogs and 20 crickets	Species classification	WASN

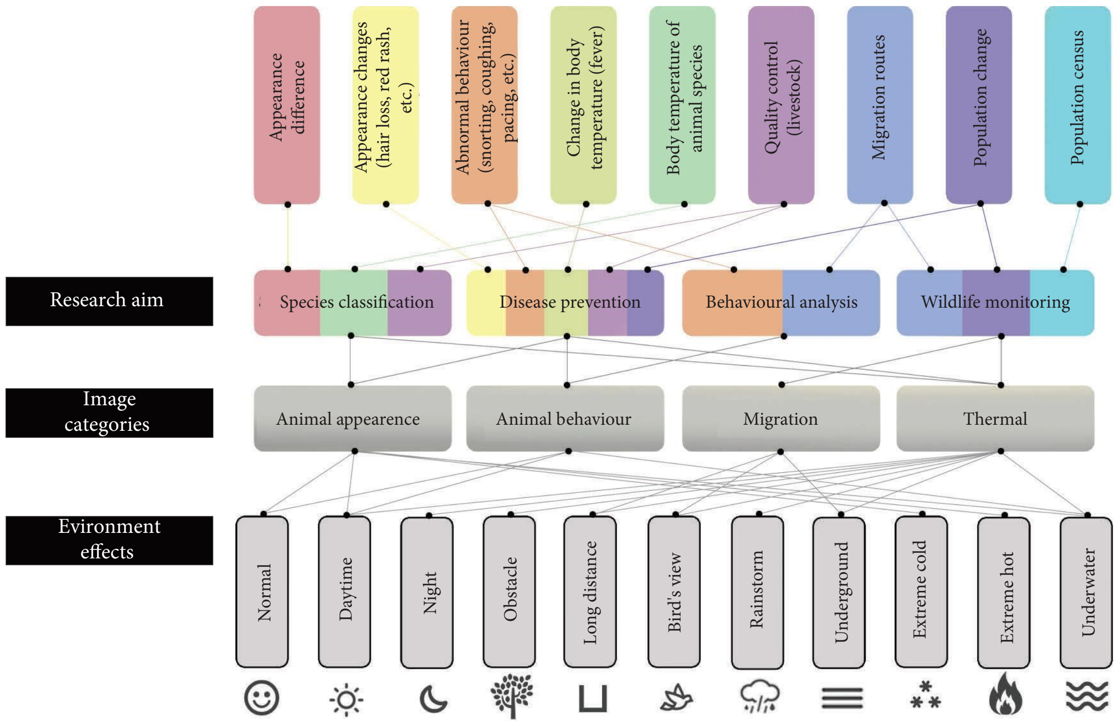


Figure 3: Relationship between research aims, four image categories and environmental effects.

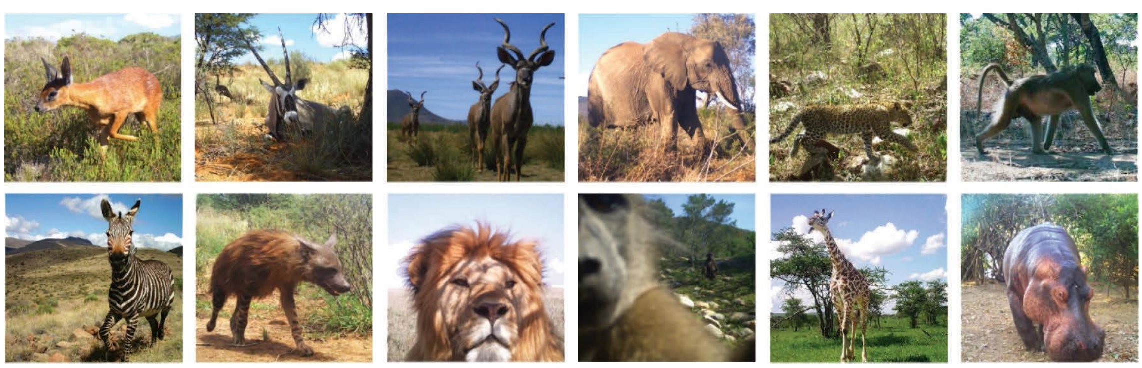


Figure 4: Samples of Snapshot Serengeti Project. Image by Swanson et al. (licensed under https://creativecommons.org/licenses/by/4.0/). This figure shows several animal images captured by camera traps located at Serengeti National Park, Tanzania [86].

4.1.2.2. Behaviour. The use of automated animal identification in modern commercial farms is gaining ground as researchers determine that animal behaviour is closely related to its health status. Okura et al. utilised images captured by an RGB-Depth (RGB-D) camera to identify individual dairy cows to detect their health conditions [41]. The standout feature of this camera is that it can capture RGB colour and depth. RGB-D images distinguish themselves from conventional RGB images by incorporating an additional depth channel alongside the standard red, green and blue channels. This depth channel conveys the spatial distance from the camera sensor to each pixel in the image, thus enabling the construction of a three-dimensional spatial representation of the scene. Figure 5 shows the visual difference between those two types of images. The advantage of this method is that it can extract the texture features of cows to identify individuals and extract gait features to complement each other when the texture is adversely affected. Specifically, gait recognition is to identify the foreground cows through a depth camera and extract gait features from aligned 3D point clouds within a walking cycle. Ultimately, they achieved rank-1 recognition accuracy (over 75%) in both normal and textureless scenes (night or those of cows with textureless coats) based on a combination of both recognition algorithms.

4.1.2.3. Migration. Image recognition technology is also useful when recognising large number of animals. As a cheaper, safer and more suitable image acquisition method, unmanned aerial vehicles (UAVs) have gradually replaced light aircraft in investigating large mammals in recent years, especially when equipped with high-precision sensors that can obtain high-quality images.

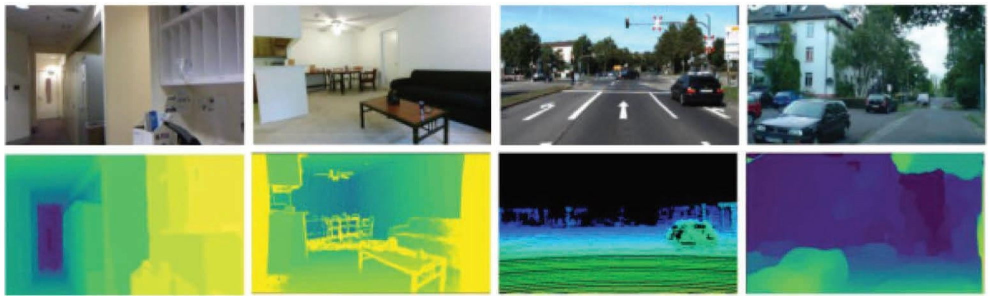


Figure 5: Samples of RGB images and RGB-D images. Image by Lopes, Souza and Pedrini (licensed under https://creativecommons.org/licenses/by/4.0/). This figure shows several samples of RGB images (first row) and their RGB-D images (second row) [87].

Delplanque et al. evaluated three multiclass CNN algorithms, Faster Region–based Convolutional Neural Network (Faster R-CNN), Libra Region–based Convolutional Neural Network (Libra R-CNN) and RetinaNet, to identify six mammals from aerial images [42]. The Libra-R-CNN outperforms the other two on the test set and even performs well in detecting a few species. The species with lower identification are caused by a variety of factors, including

1. Overlapping animal bounding boxes created during inference.
2. Different times when aerial images were taken, leading to greater variation in shadows, colours and brightness.
3. The different scenes in the dataset (trees, grass, rocks, etc.) [88].
4. Fewer learning samples.

4.1.2.4. Thermal. TR is different from image recognition in that thermal image features replace the texture features of the animal's appearance. Thermal images are better at identifying animals when the recognition rate by ordinary visual spectrum images is low in some specific conditions.

Identification of large mammals (> 350 kg) or medium-sized mammals (15–350 kg) is easy because they generally emit more heat, resulting in more visible contours on thermal images. However, recognising animals with small sizes (< 15 kg) presents significant challenges as they are usually only a few pixels on thermal images [72]. This is due to the aerial photography equipment being kept far from the ground to minimise the impact on wildlife.

Ulhaq et al. proposed a species detection system using transfer learning and enhanced small object detection algorithms [72]. The distant object detection algorithm, called Distant-YOLO (D-YOLO), was trained on a thermal imaging dataset consisting of rabbits, feral pigs and kangaroos. One of the problems with traditional CNNs is that they cannot handle low-resolution images. So, Ulhaq et al. integrated dilated convolution in the YOLO architecture to increase its receptive field for handling small objects. They achieved reasonably good recognition rates for three animals: rabbit = 98.17%, pig = 97.34% and kangaroo = 99.48% [72].

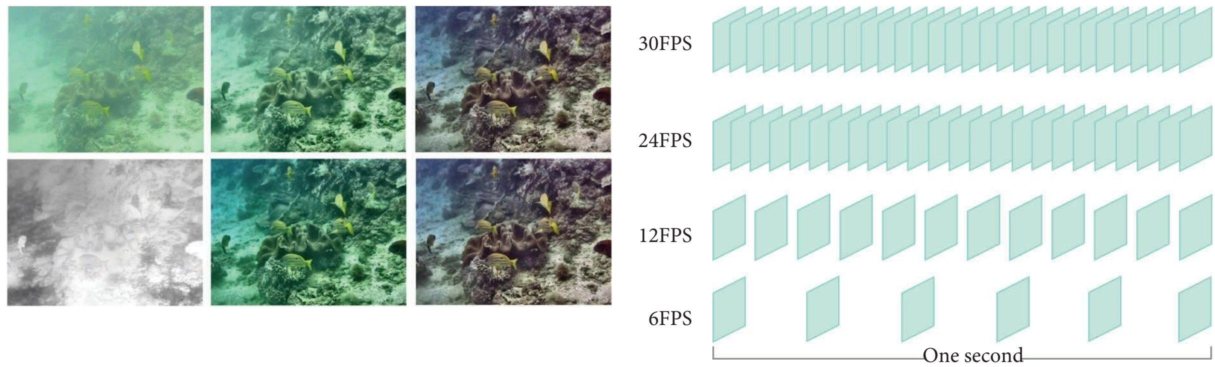
4.1.3. Video. The boundaries between video and image processing are blurred because the features they extract are all based on animal appearance. The difference is that video recognition is used to parse dynamic images, emphasising the connections between frames. These dynamic images or continuous frames contain information that cannot be contained in a single frame, such as changes in animal behaviour patterns and migratory routes.

Guzman-Pando and Chacon-Murguia proposed a novel model consisting of two CNNs to recognise moving objects in videos [61]. First, the Peripheral-CNN simulates peripheral vision to detect general motion in the scene. Second, the DeepFoveaNet detects small moving objects that cannot be captured by the former by generating a magnification of an area of interest. After evaluating 53 videos, DeepFoveaNet shows robustness in different video scenes and maintains high performance. It does not rely on previously trained neural networks, nor does it depend on a large number of training images for training, which provides new directions for future artificial vision systems.

However, this technique also has some drawbacks. Zhao et al. identified underwater creatures from 20 low-resolution videos (five of them at 640×480 pixels and 15 at 320×240 pixels) [81]. They pointed out that low-quality video resulted in the loss of fish texture and increased motion blur when swimming. It may be further exacerbated in low-light situations. Figure 6(a) shows the original blurred underwater photo and some images with enhancement algorithms.

The easiest way to solve this problem is to increase the resolution. Figure 6(b) presents standard video frames such as 6, 12, 24 and 30 frames, which means the number of pictures corresponding to one second [89]. The number of frames per second can vary depending on the requirements. As the resolution increases, the size of the video grows exponentially. It brings massive pressure on local information collection equipment or the throughput of the servers in the cloud.

Chen et al. applied two trained CNNs into a video clip directly in their study [38]. They handled a dynamic video as a continuous single-frame image. They took greyscale images and only detected frames of interest to improve recognition speed. Also, they applied a dynamic threshold process to avoid false positives caused by small image changes. This is like converting acoustic signals into



(a) (b)

Figure 6: Different video frame and restoration algorithms. (a) Original underwater image and different restoration algorithms. Image by Zhao et al. (licensed under https://creativecommons.org/licenses/by/4.0/) [81]. (b) One second of animation.

spectrograms and using deep learning to identify animals [50]. In fact, these methods ultimately rely on image processing.

Figure 7 depicts the relationship between image processing and other signal-processing technologies used in the selected 54 publications. The size of each bubble represents the number of publications contained within it. The three small bubbles relate to image processing systems. Some publications utilise two methods to obtain the required information; one is related to all three categories. This outcome is predicated on various factors, such as neural networks being more powerful and efficient in image processing and image processing time and equipment costs being lower. Further details are explained in Section 4.2.

4.1.4. Wireless Signal. Wireless communication generally refers to the transmission of information between two or more points that do not use electrical conductors as the medium [90]. It played an important role in the early field of animal recognition, which is mentioned in Section 2.2. Wireless signal technology has often been used as part of data collection rather than at the core of animal identification in recent years of research due to its cost-effective and easy-to-operate features. For example, in their research, Høye et al. proposed using vertical-looking radars (VLRs) to detect the flight trajectory and body shape of insects [69]. Wireless communication technology is irreplaceable in real-time tracking. Thangavel and Shokkalingam used GPS to detect elephants' location in real time and monitor their travel trajectories to protect nearby villagers from harm [70].

4.2. Machine Learning for Animal Recognition. According to the analysis of the 54 publications in the previous section, most recognition research is related to machine learning. As a branch of AI, machine learning has gradually developed into a multidomain interdisciplinary area in the past few decades, especially in the work that replaces humans to perform many repetitive operations [91]. This section presents the contributions of machine learning to animal recognition.

4.2.1. Open-Source Databases. Data training is fundamental and widely used in machine learning. Training data for supervised and semisupervised machine learning requires high-quality labelled datasets. On the one hand, this consumes a lot of time, and the marking (recognition and identification) of animals is a highly specialised field; this requires professionals to assist in marking. On the other hand, while not requiring labelling, unsupervised learning may also be expensive for relevant datasets [92]. The good news is that open-source databases provide the material for training and experimentation for researchers with a small budget or insufficient time to collect data. Some classical databases, such as Microsoft COCO [93], MNIST database [94] and SS Project [86], have been used extensively for this purpose. When project datasets intersect with these large datasets, utilising these datasets as a supplement for training can improve the system's performance.

Among these datasets, the ImageNet project provides researchers with more than 20,000 categories, 14 million hand-annotated images and at least 1 million bounding boxes [95]. Since some popular neural network architectures contain more than a few million parameters, some researchers use a model pretrained on a large dataset such as ImageNet and reuse it on their own collected dataset [32, 66]. Some researchers fine-tune their system on this basis to make the model perform better on their own datasets [39].

Similarly, Ferreira et al. used part of the training data from the MS COCO database in their study of three kinds of birds [55]. Since manual labelling is time-consuming, they first decided to train with 200 labelled images for 10 epochs and observed the performance. If the desired result was not achieved, they continued to add marked images until satisfied. This method saves resources.

Some researchers have also provided smaller-scale image datasets. These datasets often include categories not covered by larger datasets. For instance, Zhang et al. have curated a dataset consisting of 47 species observed in Australia and have provided manual annotations [96]. These classes, categorised by their scientific names, often encompass finer-grained features, thereby offering more valuable resources for animal image classification research.

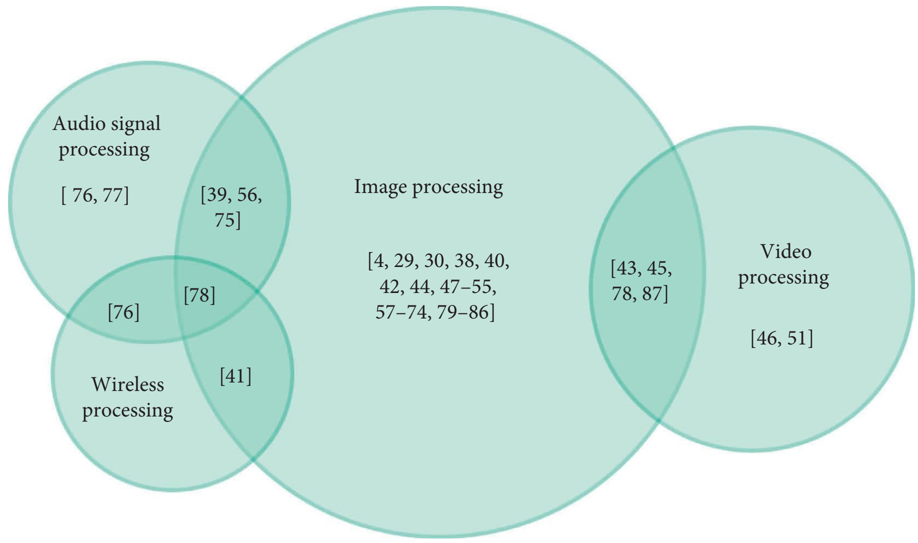


Figure 7: The relationship between image processing and other signal processing technologies in the 54 publications.

However, public datasets are not perfect, and they also have drawbacks. Psota mentioned in a study on pig recognition that the lack of some specific targets in open-source datasets makes researchers face many challenges [43]. For these species-specific studies, researchers still need to spend time collecting data and manually labelling them.

4.2.2. Deep Learning Architecture. There are many types of neural networks in deep learning, such as CNNs, recurrent neural networks, DBNs and deep reinforcement learning. They have different functions and have been widely used in various fields. The differences between these different neural network models mainly lie in the activation rules of neurons, the neural network model's topology and the learning algorithm's parameters. This section will discuss the applications of animal identification. While building a model from scratch may have better performance, it also consumes a lot of time and effort. CNNs are applied in many computer vision projects. A traditional CNN consists of some convolutional layers and some fully connected layers.

Figure 8 shows the different networks that have been used in selected publications. Figure 9 presents the number of networks used in these publications.

In some studies, the researchers tested the model with only one kind of neural network. For example, Shepley et al. developed a deep learning model to identify the three categories of animals—rhino, striped hyena and pig [60]. Their datasets are derived from photos obtained via Python scripts from several large photo websites. These photos are all captured through camera traps, including images from different periods and places. They then used a RetinaNet model to train target animal pictures on three datasets. This one-stage detector introduces Feature Pyramid Networks (FPNs) and Focal Loss. The RetinaNet architecture consists of four components: bottom-up pathway as a backbone network, top-down pathway, lateral connections to sample coarse feature maps and two subnetworks which are classification subnetwork and regression subnetwork [97]. They point out that RetinaNet is used to strike a balance between computational efficiency and accuracy. In other words, the choice of model is based on the purpose of the research.

Singh and Mumbarekar presented a neural network model to classify benthic animals, using a classifier and a detector [62]. They used the Inception V3 model, a 48-layer DNN, to build the classifier and overcome the overfitting problem. Inception V3 has fewer than 25 million parameters compared to the 60 million parameters of the popular CNN, the AlexNet [98]. For the detector, they combined depth-wise convolution and point-wise convolution to optimise computations in a mobile-net architecture and used Single Shot Multibox Detector (SSD) to improve the speed of feature extraction. Additionally, they supplemented the dataset with complex background images, addressing the limitations of images captured under an electron microscope, allowing machine learning models to recognise objects in interfering environments.

Some studies train multiple networks simultaneously time to select the network that can achieve satisfactory accuracy. Simonyan and Zisserman used two different architectures, VGG-16 [99] and AlexNet [98], on a dataset of 20 species which include animals and humans, and observed their performance [66]. Those two popular architectures are all based on CNN and are also suggested for the ImageNet Large Scale Visual Recognition Challenge (ILSVRV) [100] to classify images. VGG-16 is an improvement on AlexNet. The first five layers of AlexNet are convolutional, while VGG-16 is composed of five groups of convolutional layers, each with 3–5 convolutional layers. It means that this model is deeper and broader than the traditional CNN model. The training time will be longer, and the trained model size will exceed 500 megabytes. The testing results show that AlexNet provides about 90% accuracy and VGG-16 provides about 93% accuracy.

By comparing a selection of 54 publications, the top seven network models in terms of utilisation are listed. The

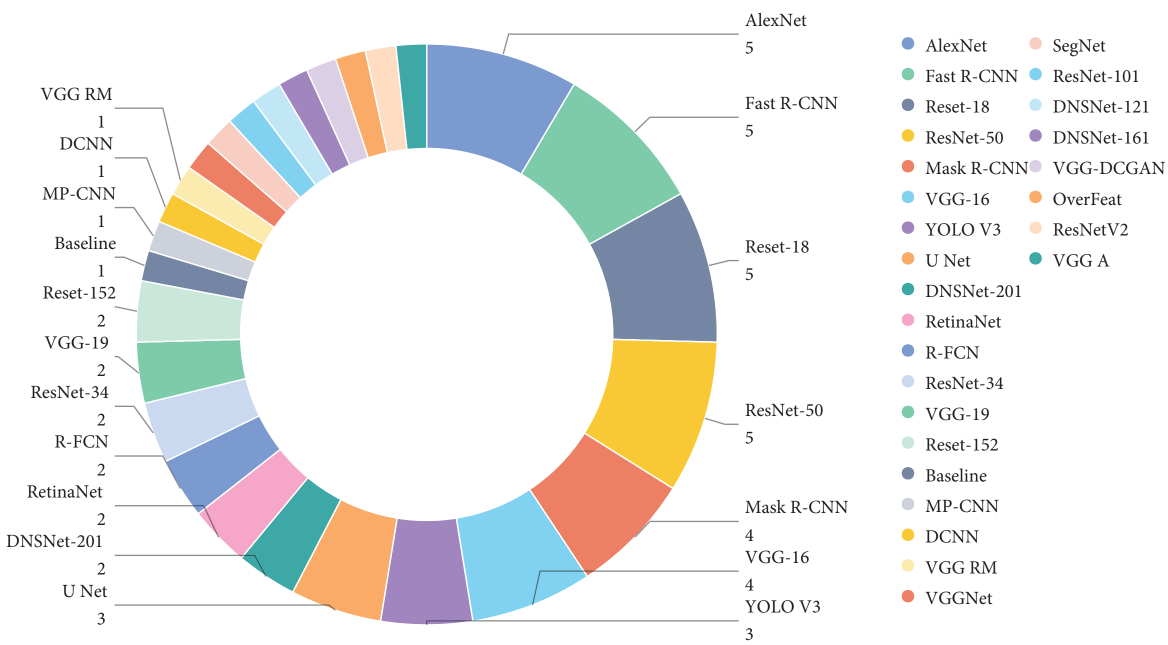


FIGURE 8: Summary of networks that have been used in publications.

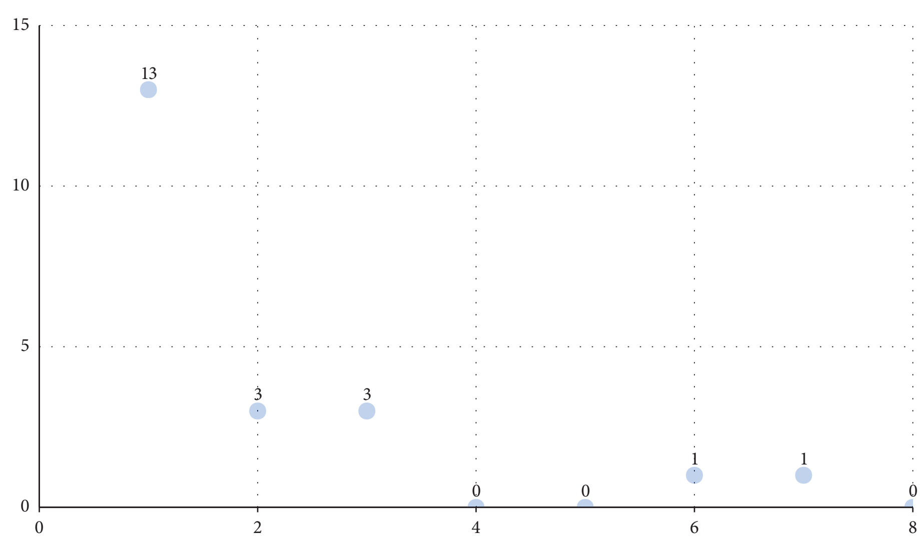


FIGURE 9: The number of networks that have been used in every publication.

number of species they identify and their accuracy rates are summarised, which can be found in supplementary table 1.

4.2.3. Machine Learning Approaches. Machine learning approaches can be classified into three classic categories based on the algorithm used in training: supervised learning, unsupervised learning and reinforcement learning [101]. This section will discuss the different approaches applied in the selected publications.

4.2.3.1. Supervised Learning. In supervised learning, each example is a pair consisting of a labelled input object and the desired output value [102]. In general, computers learn through a set of labelled datasets. For accurate prediction or classification, the input data are marked as the correct answer. Most computer vision systems used for classification purposes are based on supervised learning. This process is time-consuming and labour-intensive and sometimes requires the assistance of professional researchers.

An animal census study based on a single CNN architecture has been proposed by Kellenberger, Marcos and Tuia [73]. Their dataset is aerial imagery captured using UAVs. Before training, the researchers needed to label the photos manually. Although the total number of images was only 654, it took the researchers three days for labelling and another three days for validation.

So much time commitment is not viable for large datasets. Some studies use classifiers to assist in detecting regions of interest. Marsot et al. used the Haar Cascade classifier trained by the OpenCV library to detect pig faces [22]. First, the dataset used to train the classifier consists of 564 positive samples and 2110 negative samples. Each positive sample generates a file containing the file name and face information through the opencv_create samples function, while each negative sample is a picture without pig face information. Then a second classifier is used to detect eyes, and a training process similar to the first one is used. To avoid false positives, geometric constraints and shallow convolutional networks are used. The final false positive rate of 5.6% was deemed satisfactory, but the researchers considered their dataset as insufficient. With sufficient datasets, the false positive rate can be lowered.

Interestingly, automated annotation was done using existing equipment in a study of birds. Ferreira et al. linked a Raspberry Pi with an RFID data reader and a Pi camera and placed them near an artificial feeder [55]. When a bird with a PIT tag approaches, the RFID recorder will transmit the data of the PIT tag to the Raspberry Pi and turns on the camera to take one picture every 2 s. To complete the labelling process, the saved photos were automatically named as the birds' names. The disadvantage of this method is that when multiple birds stay on the feeder, it is impossible to tell which bird the photo belongs to, resulting in label errors.

4.2.3.2. Semisupervised Learning. In semisupervised learning, only part of the training data is labelled and the rest is not. Meena Agilandeeswari proposed a two-stage animal breed classification model [48]. In the first stage, they trained the TensorFlow version of Inception V3 on the ImageNet dataset. To deal with misclassification caused by unlabelled data, they adopted 'pseudo-labelling based on semi-supervised learning.' Specifically, after the supervised data training, researchers tried labelling the unlabelled data and linking them with the original labels. The second stage uses Multipart Convolutional Neural Network (MP-CNN) for Fine-Grained Classification (FGC). Among them, they propose a hybrid feature extraction framework based on Fisher Vector (FV) and Stacked Autoencoder (SAE) ensembles to extract animal features. The advantage of this algorithm is that it can handle nonlinear information loss, thereby extracting part features. Experimental results demonstrate that their method achieves higher accuracy and less processing time compared to a single DNN architecture.

4.2.3.3. Unsupervised Learning. Unsupervised learning does not contain labels. It outputs based on the trend and commonality of the input. Clustering is one of the most common unsupervised learning methods.

Some studies use unsupervised learning in animal classification as well. Tan, Zhou and He proposed a model based on FCM-KM and Mask R-CNN to identify five kinds of butterflies [68]. *K*-Means is an unsupervised learning algorithm. If the selected *K* value is not reasonable, it will seriously affect the clustering accuracy. Therefore, the authors proposed to use the *K*-Means clustering algorithm optimised by the firefly algorithm and the max-min distance algorithm to determine the number *K*. In addition, they also compared the traditional Mask R-CNN, and the results show that the accuracy of the improved model is higher than that of the traditional model.

4.3. Solutions for Processing Large-Scale Datasets in Animal Classification. To relieve humans of repetitive work and improve system efficiency, machine learning is being deployed to process large amounts of data. However, a thorough review of the publications found that the application of machine learning solutions cannot be generalised. Often, researchers need to tailor a unique plan to suit the application domain.

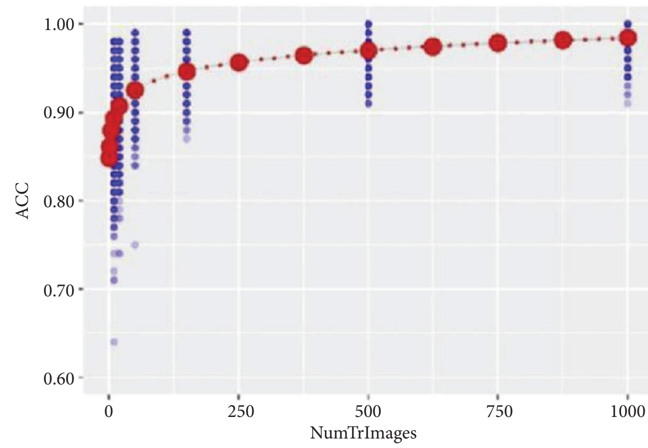
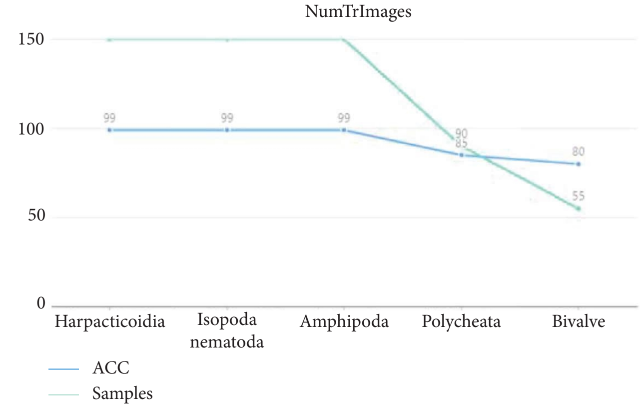
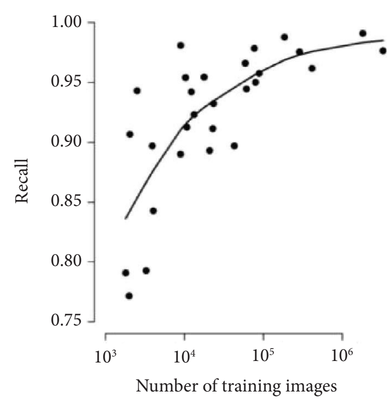
4.3.1. Data Augmentation and Imbalanced Datasets. Deep learning systems that exploit large amounts of data are primarily developed outside the field of ecology. Shahinfar, Meek and Falzon designed seven training sets with different numbers of samples (10, 20, 50, 150, 500 and 1000) [58]. They used six variants of Residual Neural Network (ResNet) and Densely Connected Neural Network (DenseNet) to perform transfer learning based on ImageNet. The results were as they predicted, increasing accuracy with more images in the training set. Generally, the performance improvement raised rapidly initially, slowed from 150 images onwards, with no significant improvement after 500 images.

Research on benthic animals shows a similar result [62]. Despite impressive recognition rates for some species (99%), the model requires a high sample size. The Polychaeta family with 90 samples had a recognition rate of 85%, while the Bivalvia family with 55 samples was only 80% accurate.

Another study with a large number of data and species highlighted the impact of imbalanced datasets on accuracy and the benefits of data augmentation. Tabak et al. used TensorFlow to train a deep CNN (ResNet-18) to classify wild animals from their photos [64]. Their dataset included 3,367,383 photos captured by camera traps at five locations in the United States for a total of 27 animal species. Among them, the species containing the most and the fewest samples differed by a factor of nearly 1000. Therefore, they used data augmentation to compensate for the drawbacks brought about by the imbalanced dataset to make the model more robust. The trained model can classify approximately 2000 images per minute on a laptop with 16 GB RAM. Their test results included 10% of the American animal species datasets with a recognition rate of 98%, and the Canadian animal recognition rate was 82%. In addition, the model's performance in identifying new datasets consisting of untrained animals and empty sets also provides ideas for preliminary classification of future images. Table 5 illustrates the relationship between the number of samples and accuracy in the above three studies.

4.3.2. Joint CNN. A study on the Wildlife Spotter Dataset of Southcentral Victoria containing 125,621 samples discusses a Deep CNN framework [39]. Margarita and Andrey

Table 5: Relationship between number of samples and accuracy in different studies.

No. of samples	Accuracy and number of samples	Ref.
10, 20, 50, 150, 500, 1000		[58]
150, 150, 150, 90, 55		[62]
1, 804-3, 367, 365		[64]

proposed a joint CNN architecture that includes multiple subsets. These use two VGG-16 branches as a partial feature extractor network and one VGG-19 for overall shape recognition. The advantages of this joint architecture are worth considering further. When the three branches have the right conditions to work at the same time, the recognition accuracy becomes close to 100%. However, a disadvantage is also present. It requires long-term learning more than other architectures, including enough datasets to support individual learning for each branch. Unbalanced datasets, photos of forest sites with blurred backgrounds and small birds burden the architecture and lower the recognition rate.

4.3.3. Pretraining and Transfer Learning. Moallem et al. provided an animal classification and detection system based on the pretrained Xception architecture, which considered some specific factors of the Texas Parks and Wildlife Department (TPWD) dataset [4]. They pretrained two Xception architectures on ImageNet, a large object classification dataset. Notably, the researchers found data drift by observing images of the site in different weather conditions and during all four seasons. Consequently, they proposed an Automatic Retraining Procedure, which can be used to judge whether retraining is needed based on the Retraining Trigger Index (RTI). This procedure holds significance for field sampling and local screening, which ensures model robustness through continuous training, mitigating the impact of background structure changes over time due to light and vegetation variations. Consequently, the accuracy of recognition can be kept high while obtaining a large number of field photos.

4.4. Limitations of Animal Classification in Diverse Applications. Automatic identification of some specific animals has been used in modern farms and it has achieved reasonable results [22, 41, 43, 49]. However, most studies about animal species classification are still small-scale experiments. In fact, researchers have already pointed out the importance of tracking species diversity, and due to a series of climate and environmental changes in recent years, this issue has attracted greater attention in research circles.

4.4.1. Image Quality and Environmental Impact. The application of deep learning in image recognition is highly dependent on the quality of the image. Photos captured by digital cameras cannot achieve high quality under some special environmental conditions.

First, day and night photographs have considerable differences. In the absence of adequate lighting, standard digital photos capture minimal animal features [103]. Although artificial lighting can be used to reduce the environmental impact, excessive light exposure may cause anxiety in animals.

Second, the background of images is more uncontrolled in the wild environment. Unforeseen weather and geological changes, vegetation cover, animal behaviour and equipment maintenance issues are all important factors that may affect the quality of the photo [103]. Underwater photography faces challenges due to reduced light, water turbidity and background clutter. In a study on underwater biometrics, Demir, Christen and Ozev pointed out that balancing performance and power consumption is critical in changing natural environments [44]. Their proposed framework consists of a neural network recognition block that obtains single-frame confidence and an adaptive decision block. The latter can dynamically change the system parameters based on environmental conditions. Specifically, the researchers augmented images according to simulate turbidity and illumination to train data. On the other hand, the multiframe decision algorithm minimises power consumption by controlling the camera frame rate and LED lighting levels. The test results show that the model can maintain a recognition rate of about 90% and save about 92.7% of energy consumption.

In another study, the effect of background on images was more pronounced. Generally, cameras on farms or in the field are fixed to capture pictures in one position automatically. However, models trained with images from the same background may not achieve similar accuracy in another environment. Schneider et al. proposed a study with 47,279 images from 36 unique locations and trained the model [54]; the results showed a recall rate of 0.971 ± 0.0137 when testing species classification on locations used during training. However, the same species achieved only 68.7% accuracy for those other locations. From this, it can be concluded that the influence of background on species identification is significant.

To deal with issues caused by the background, some studies tested the following solutions. Fang et al. proposed a network to segment background from objects of interest called Tiramisu [53]. This kind of network can segment the photos in the dataset before identification or classification to remove the influence of background noise. These researchers selected 10 categories containing animals, plants and objects from the Microsoft COCO dataset [93], each containing 1000 images. To improve network performance, they reduced the number of layers in Dense Block without affecting the experimental results. They selected three CNN models to test the photos before and after background removal in the classification stage. The results show that the image recognition accuracy after background removal is slightly higher than that of the unprocessed image. In addition, for those pictures that cannot be recognised before processing, the recognition rate after background removal reached 34%–73% accuracy, for different networks.

4.4.2. Data Sources and Feasibility. The training phase for machine learning requires a large amount of data to support it. According to estimates, about 8.7 million species (±1.3 million) are on the earth [104], of which 6.5 million are on land, and 2.2 million live in the deep ocean. Some specific organisms, due to factors such as their size and behavioural characteristics, are difficult to capture using automated field capture devices or traditional digital cameras, resulting in a scarcity of training data. While data augmentation can help reduce overfitting, the recognition rates are still not comparable to species with sufficient sample sizes.

Hoye et al. noted that although insects represent the vast majority of animal species, the lack of data applies especially to insects [69]. The excellent performance of deep learning in other fields has not yet been applied to invertebrate monitoring. They proposed a variety of ways to collect insects' biological information, such as camera traps designed with light sources or natural resource traps, VLRs to capture information such as the trajectory, speed and body shape of flying insects and the use of CNNs to identify wing beats and the acoustic signal produced by their movement. However, these methods have only been used in the laboratory, and there are insufficient data to support whether these methods can work in the field environment.

Building a dataset that counts all or most species to identify different species is impractical for an individual group or organisation. Even though there are excellent algorithms to optimise the recognition process, comparing different species in a vast database can still require excessive time for the recognition process.

4.5. The Significance of Taxonomic Rank in Animal Identification. Whether it is the traditional intrusive labelling method or the computer image technology with neural networks as the animal recognition method, classification is based on an individual or a ‘species’. In other words, the methods mentioned above treat the animal as a specific ‘object’ and ignore the evolutionary chain and the relationship between species. After analysing the existing research, this section will explain the significance of taxonomic rank in identifying animal species.

4.5.1. Taxonomic Impediment. Taxonomy, a critical scientific methodology for grouping, naming, defining and classifying organisms into specific communities, is essential for biodiversity conservation. The Convention on Biological Diversity report outlines challenges in taxonomy development, emphasising the need for training taxonomy experts and strengthening infrastructural support [105]. The Darwin Conference in 1998 further underscored the need to develop taxonomies that aid biodiversity efforts, addressing classification barriers like knowledge gaps and professional shortages. Furthermore, the traditional taxonomy process, requiring complex laboratory equipment and significant time investment from biologists, underscores the necessity for a more systematic and efficient identification system.

4.5.2. Species Evolution and Discovery of New Species. Taxonomic rank refers to a scientific grading system in biology. It includes eight major ranks: domain, kingdom, phylum, class, order, family, genera and species. It also includes some finer subranks.

Figure 10 shows the taxonomic rank of the Australian eastern grey kangaroo. First, its scientific name, which is also its species, is *Macropus giganteus*. In the next rank above, the genus *Macropus*, only two species are left: western grey kangaroo and eastern grey kangaroo. The family Macropodidae is their closest relative, including some other animals with long feet, commonly known as kangaroos, quokkas and wallaroos. The order Diprotodontia, which means ‘two forward teeth’, is the next higher rank, which includes koalas, wombats and many herbivores. Next comes the Mammalia class that nourishes the next generation by breastfeeding, which is classified among the animals with a backbone and bilaterally symmetric in the phylum Chordate. Finally, the animal kingdom is represented by all multicellular organisms classified as the Eukarya domain, comprising organisms with eukaryotic cells found in plants and fungal cells.

The contemporary taxonomy, grounded in Carl Linnaeus’ research, focuses on evolutionary relationships among organisms, although these relationships are not absolute due to uncharted or extinct species [107]. Advancements in DNA technology and research into the microscopic world continue to refine taxonomic ranks, highlighting their critical role in taxonomic studies. Taxonomy Australia estimates indicate that describing Australia’s undescribed species could take over four centuries at the current pace. However, a revised plan aims to accelerate this process to 25 years, increasing the discovery rate by 16 times. Australian taxonomists are exploring machine learning for taxonomy, though specific protocols for animals are yet to be established.

5. Some Recent Trends

This SLR covers the publications from 2016. For an updated perspective, this chapter discusses recent publications.

Zheng et al. proposed a method to monitor changes in animal health by detecting body temperature using infrared thermography (IRT) imaging technology [108]. This noncontact information collection method has been widely used in animal husbandry being convenient and rapid to deploy. Some researchers have proposed that IRT imaging can detect temperature changes in different parts of the animals, which has important implications for the early detection of animal disease, pregnancy and stress responses. The downside, however, is that IRT is highly dependent on expensive sensors and other hardware devices, and its operators require considerable expertise. The higher the complexity of the environment, the more significant the impact is on animal body temperature. Researchers can minimise environmental impact in a farm setting by adjusting some parameters. However, this is almost impossible in the wild.

Wu et al. proposed a system for recognising multiple animal calls in a noisy environment [109]. They pointed out that the current acoustic signal recognition studies usually use a single animal voice with less background influence as training data, which leads to an unsatisfactory recognition rate when testing in the natural environment. Apart from converting acoustic signals into spectrograms, they also developed a method called ‘rainbow mapping’. Specifically, each band uses a unique gradient colour to keep frequency features. This preserves more information converted from the audio signal than a greyscale spectrogram.

More research is done on image recognition based on deep learning. Xiao et al. proposed a method to identify individual dairy cows from their top-view image [110]. They improved the Mask R-CNN model to segment the image, obtain cow shape and back pattern features and train a support vector machine (SVM) classifier. They changed the backbone network structure and added an enhanced feature fusion network after FPN processing. The model finally achieved an impressive accuracy of 98.67% on a dataset of 48 cows totalling 12,000 images.

6. Future Challenges

Publications from 2016 show that research in animal identification is increasing yearly, and people are more interested in the application of animal identification. This study shows that many of these new methodologies of animal recognition are still experimental, and some aspects of these systems can be improved. This section proposes some ideas for future investigations that can be helpful for researchers.

6.1. More Awareness of Biodiversity. Global biodiversity is facing challenges. It is urgent to arouse people’s awareness of protecting the environment and biodiversity. One of the most controversial topics in recent years has been global warming. Continuous high temperatures, melting glaciers and forest

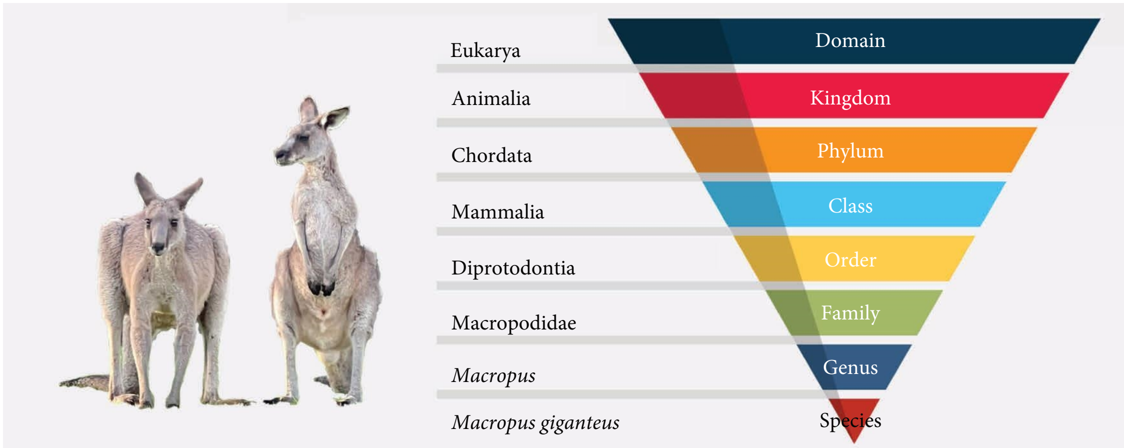


Figure 10: Taxonomic rank of eastern grey kangaroo. *Macropus giganteus* Shaw, 1790, observed in Australia by Orlovich (licensed under https://creativecommons.org/licenses/by/4.0/) [106], background removed.

fires in many countries have caused not only huge economic losses to human society but also caused many animals to lose their habitats and even die. In addition, although the frequency of poaching has been reduced in recent years with improvement in people's awareness of wildlife protection, it has not been completely eliminated. Section 4.3.1 mentioned imbalanced datasets, which indicates that some species are declining, making it difficult for the camera trap to capture their images. The reasons for the sharp decline in biodiversity are complex, but raising people's awareness of it has been a long-term goal for humans.

6.2. Challenges of Few Sample Learning. In contrast to human capability for recognising new objects and deriving conclusions from a few examples, advanced machine learning algorithms necessitate substantial data for effective training. Few-shot learning (FSL) and zero-shot learning (ZSL) are hot topics in the field of machine learning, as almost all machine learning researchers aspire to create systems that can learn and generalise from limited samples.

The status quo of biological populations cannot be changed in a short period of time, and humans cannot interfere with wild animals to keep their numbers constant. The challenge researchers face is achieving satisfactory accuracy through machine learning even without data and avoiding overfitting. Section 4 mentions that some projects suffer from imbalanced datasets and cannot achieve satisfactory accuracy, even if image augmentation techniques or algorithms have been applied to optimise the outcome with fewer samples. In a survey, Lu et al. pointed out three implications of FSL: reducing the cost of large-scale training data, bridging the gap between human intelligence and AI and quick model deployment for emerging tasks [111]. These three points are significant for the current fast-paced production applications and the long-term development of AI techniques.

6.3. More Methods to Identify Animals. As an important aspect of animal identification, it is essential to explore how various animal features, such as fur texture, ear shape, facial features, iris patterns, and body shape, can be utilized in recognition applications. However, these features may vary slightly over time, leading to potential misidentification. Some biological information databases have begun to be established because identifying species by collecting organism-specific DNA fragments and comparing them with a reference library is more accurate than other identification methods. And the samples collected are not limited to living organisms; any traces related to animals (faeces, tooth marks, secretion tracks and cell debris) may be identified [25]. This is significant, especially for those endangered species. However, this method has some disadvantages too. DNA barcode identification usually requires extracting fragments from high-quality DNA, and the identifiable DNA fragments may vary between species [25]. This means that establishing the database in the early stages requires a lot of labour and financial resources, and the identification tasks in the later stage also require exceptional equipment support.

6.4. Comprehensive Database Sharing. It was mentioned in Section 4.2.1 that the basis of machine learning is the training with a large amount of data, and most research on animal identification methods follows supervised learning. In other words, the dataset is one of the most important steps for animal identification, including the quantity of labelled targets. At present, there are some open-source datasets for different regions and different animal classes. Those datasets make excellent contributions to research in this area, especially the world's largest wildlife dataset—SS dataset [86]. This dataset contains over 7.1 million images from camera traps of Serengeti National Park in Tanzania since 2010 and is available to animal researchers for free. Professionals and nonprofessionals from around the globe have volunteered to label 61 categories and add bounding box annotations to these photos. However, some researchers have pointed out that most teams cannot recruit a large number of volunteer teams to the SS project to annotate information of interest [69]. In addition, the SS project currently only includes animals in the Serengeti National Park and no other areas. The dilemma brought by the interdisciplinary nature of this work is also one of the problems animal identification researchers face. Zoology and computer science are two different fields. It requires researchers to cooperate and spend time learning knowledge in each other's areas. Vidal et al. have highlighted the importance of cooperation between biologists and computer scientists [20].

7. Conclusion

This survey presents a comprehensive SLR on animal identification based on four related survey questions. Identifying some individual animals has been applied on modern farms and achieved good results. However, improving the practicality of animal identification technology and creating more market value is one of the problems researchers now face and one of the reasons why most studies are still small-scale experiments. Researchers have already pointed out the importance of tracking species diversity. Due to a series of climate and environmental changes in recent years, this issue has attracted attention in research circles. At the same time, animal welfare has also gradually received more attention. Five freedoms (freedom from thirst and hunger, freedom from discomfort, freedom from pain, injury and disease, freedom to express most normal behaviour and freedom from fear and distress) and three principles (living a natural life, being fit and healthy and being happy) proposed by Professor John Webster are used to judge animal welfare [112]. It predicated preference for noninvasive identification methods.

The large volume of photos obtained from camera traps worldwide makes taxonomists feel more under pressure to automate the recognition process. Recognition of animals is quite tricky using only image recognition techniques. Undoubtedly, the deep learning methodology of AI provides a more efficient solution for species classification. With the increase of open-source datasets and the introduction of new architectures, deep learning models have achieved dozens or even hundreds of times more efficient results than humans can achieve in classifying animals from their photos. Numerous research projects tackle the background noise challenge brought about by the wild or natural environment. The main contribution of this paper is to provide an overview of the current research in the field of animal identification by browsing the literature to analyse the advantages and disadvantages of various identification methods—especially deep learning, which has been the shining star of computer vision in recent years. Researchers are gradually discovering its unlimited potential, especially with the increasing performance of the GPUs. However, it is important to note that deep learning techniques also face significant limitations and challenges, such as the need for large amounts of labelled data, the complexity of training models and the potential for overfitting specific datasets. Additionally, the environmental impact of the high computational power required for deep learning cannot be overlooked.

Privacy concerns for wildlife and the impact of invasive identification techniques on animal behaviour are critical considerations. Invasive methods can cause significant stress and behavioural changes, disrupting natural activities and well-being. These methods may result in unreliable scientific data due to altered behaviour caused by fear and pain. Therefore, advancing noninvasive identification techniques is essential to minimise interference, preserve animal welfare and ensure accurate and ethical research. Noninvasive techniques, while less disruptive, still pose challenges in terms of accuracy and reliability.

Through an in-depth study of the selected literature, possible improvements to existing methodologies are explored, and some ideas for future biometric data extraction and the design of accelerated neural network training methods are provided. It is worth noting that further research needs to focus on deep learning, which is not limited to recognising species in a particular specific geographical area but is an essential part of monitoring and protecting global biodiversity. Future studies should also explore the ethical implications and develop guidelines to minimise the adverse effects on wildlife, ensuring that the advancements in technology contribute positively to conservation efforts without compromising animal welfare.

Data Availability Statement

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Qianqian Zhang: conceptualisation, data curation, formal analysis, investigation, methodology and writing–original draft preparation. Khandakar Ahmed: conceptualisation, methodology, project administration, resources, supervision and writing–review and editing. Nalin Sharda: writing–review and editing. Hua Wang: writing–review and editing.

Funding

This project has received no financial support from any funding agencies in the public, commercial or not-for-profit sectors.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. *(Supporting Information)*
The Supporting Information includes a table that provides a comparative analysis of 54 selected publications, highlighting the top seven most utilised network models. The table summarises the number of species each model can identify, along with their respective accuracy rates.

Acknowledgments

First and foremost, I express my deepest gratitude to my supervisor Dr Khandakar Amend, for his continuous support. I would not have completed my work without his professional mentoring, patient guidance, friendly communication and encouragement. I also would like to thank my co-supervisors, Dr Nalin Sharda and Professor Hua Wang, for their kind support and guidance.

References

[1] B. Fontaine, K. van Achterberg, M. A. Alonso-Zarazaga, et al., "New Species in the Old World: Europe as a Frontier in Biodiversity Exploration, a Test Bed for 21st Century Taxonomy," *PLoS One* 7, no. 5 (2012): e36881–e36887, https://doi.org/10.1371/journal.pone.0036881.

[2] M. Costello and S. Wilson, "Predicting the Number of Known and Unknown Species in European Seas Using Rates of Description," *Global Ecology and Biogeography* 20, no. 2 (2010): 319–330, https://doi.org/10.1111/j.1466-8238.2010.00603.x.

[3] W. Reid, H. Mooney, A. Cropper, D. Capistrano, S. Carpenter, and K. Chopra, *Millennium Ecosystem Assessment. Ecosystems and Human Well-Being: Synthesis* (Washington, D.C: Island Press, 2005).

[4] G. Moallem, D. D. Pathirage, J. Reznick, J. Gallagher, and H. Sari-Sarraf, "An Explainable Deep Vision System for Animal Classification and Detection in Trail-Camera Images With Automatic Post-Deployment Retraining," *Knowledge-Based Systems* 216 (2021): 106815, https://doi.org/10.1016/j.knosys.2021.106815.

[5] T. G. B. I. Facility, "Catalogue of Life Checklist: Chordata3," (2022), https://www.gbif.org/species/44.

[6] B. Kitchenham, *Procedures for Performing Systematic Reviews* (Keele, UK: Keele University, 2004).

[7] J. Blancou, "A History of the Traceability of Animals and Animal Products," *Revue Scientifique et Technique (International Office of Epizootics)* 202, no. 2 (2001): 413–425, https://doi.org/10.20506/RST.20.2.1282.

[8] T. M. Krier, "Horse Breeding in the Medieval World," *Medium Aevum* 68, no. 1 (1999): 184.

[9] K. S. Schwartzkopf-Genswein, J. M. Stookey, and R. Welford, "Behavior of Cattle During Hot-Iron and Freeze Branding and the Effects on Subsequent Handling Ease," *Journal of Animal Science* 75, no. 8 (1997): 2064–2072, https://doi.org/10.2527/1997.7582064x.

[10] M. Neary and A. Yager, "Methods of Livestock Identification," (2002), https://www.researchgate.net/publication/244351785_Methods_of_Livestock_Identification#:~:text=...-,Neary%20and%20Yager%2C%202002).,Neary%20and%20Yager%2C%202002)%20.

[11] G. Caja, C. Conill, R. Nehring, and O. Ribó, "Development of a Ceramic Bolus for the Permanent Electronic Identification of Sheep, Goat and Cattle," *Computers and Electronics in Agriculture* 24, no. 1-2 (1999): 45–63, https://doi.org/10.1016/S0168-1699(99)00036-8.

[12] A. S. Voulodimos, C. Z. Patrikakis, A. B. Sideridis, V. A. Ntafis, and E. M. Xylouri, "A Complete Farm Management System Based on Animal Identification Using RFID Technology," *Computers and Electronics in Agriculture* 70, no. 2 (2010): 380–388, https://doi.org/10.1016/j.compag.2009.07.009.

[13] J. W. Gibbons and K. M. Andrews, "PIT Tagging: Simple Technology at its Best," *BioScience* 54, no. 5 (2004): 447–454, https://doi.org/10.1641/0006-3568(2004)054[0447:PTSTAI]2.0.CO;2.

[14] F. Sánchez-Vizcaíno, P. J. M. Noble, P. H. Jones, et al., "Demographics of Dogs, Cats, and Rabbits Attending Veterinary Practices in Great Britain as Recorded in Their Electronic Health Records," *BMC Veterinary Research* 13, no. 1 (2017): 218–313, https://doi.org/10.1186/s12917-017-1138-9.

[15] K. Stanford, J. Stitt, J. Kellar, and T. McAllister, "Traceability in Cattle and Small Ruminants in Canada," *Revue Scientifique et Technique de l'OIE* 20, no. 2 (2001): 510–522, https://doi.org/10.20506/rst.20.2.1291.

[16] W. M. S. Russell and R. L. Burch, *The Principles of Humane Experimental Technique* (Methuen, MA: Methuen, 1959).

[17] M. J. Prescott and K. Lidster, "Improving Quality of Science through Better Animal Welfare: The NC3Rs Strategy," *Lab Animal* 46, no. 4 (2017): 152–156, https://doi.org/10.1038/laban.1217.

[18] G. Fleming, *Rabies and Hydrophobia: Their History, Nature, Causes, Symptoms, and Prevention* (London, UK: Chapman and Hall, 1872).

[19] S. Neethirajan, "Recent Advances in Wearable Sensors for Animal Health Management," *Sensing and Bio-Sensing Research* 12 (2017): 15–29, https://doi.org/10.1016/j.sbsr.2016.11.004.

[20] M. Vidal, N. Wolf, B. Rosenberg, B. P. Harris, and A. Mathis, "Perspectives on Individual Animal Identification From Biology and Computer Vision," *Integrative and Comparative Biology* 61, no. 3 (2021): 900–916, https://doi.org/10.1093/icb/icab107.

[21] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics* (Berlin, Germany: Springer Science & Business Media, 2007).

[22] M. Marsot, J. Mei, X. Shan, et al., "An Adaptive Pig Face Recognition Approach Using Convolutional Neural Networks," *Computers and Electronics in Agriculture* 173 (2020): 105386, https://doi.org/10.1016/j.compag.2020.105386.

[23] J. A. Eikelboom, J. Wind, E. van de Ven, et al., "Improving the Precision and Accuracy of Animal Population Estimates With Aerial Image Object Detection," *Methods in Ecology and Evolution* 10, no. 11 (2019): 1875–1887, https://doi.org/10.1111/2041-210X.13277.

[24] W. Disney, J. Green, K. Forsythe, J. Wiemers, and S. Weber, "Benefit-Cost Analysis of Animal Identification for Disease Prevention and Control," *Revue Scientifique et Technique de l'OIE* 20, no. 2 (2001): 385–405, https://doi.org/10.20506/rst.20.2.1277.

[25] H. C. Rees, B. C. Maddison, D. J. Middleditch, J. R. Patmore, and K. C. Gough, "Review: The Detection of Aquatic Animal Species Using Environmental DNA–A Review of eDNA as a Survey Tool in Ecology," *Journal of Applied Ecology* 51, no. 5 (2014): 1450–1459, https://doi.org/10.1111/1365-2664.12306.

[26] A. Jukan, X. Masip-Bruin, and N. Amla, "Smart Computing and Sensing Technologies for Animal Welfare: A Systematic Review," *ACM Computing Surveys* 50, no. 1 (2017): 1–27, https://doi.org/10.1145/3041960.

[27] S. Schneider, G. W. Taylor, S. Linquist, and S. C. Kremer, "Past, Present and Future Approaches Using Computer Vision for Animal Re-Identification From Camera Trap Data," *Methods in Ecology and Evolution* 10, no. 4 (2019): 461–470, https://doi.org/10.1111/2041-210X.13133.

[28] L. Hiby and P. Lovell, "Computer Aided Matching of Natural Markings: A Prototype System for Grey Seals," *Report of the International Whaling Commission* 12 (1990): 57–61.

[29] P. C. Ravoor, "Deep Learning Methods for Multi-Species Animal Re-Identification and Tracking–A Survey," *Computer Science Review* 38 (2020): 100289, https://doi.org/10.1016/j.cosrev.2020.100289.

[30] T. Petso, R. S. Jamisola, and D. Mpoeleng, "Review on Methods Used for Wildlife Species and Individual Identification," *European Journal of Wildlife Research* 68, no. 1 (2022): 3–18, https://doi.org/10.1007/s10344-021-01549-4.

[31] C. Wohlin, *Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. EASE '14* (New York, NY: Association for Computing Machinery, 2014).

[32] M.S. de Arruda, G. Spadon, J. F. Rodrigues, W. N. Gonçalves, and B. B. Machado, "Recognition of Endangered Pantanal Animal Species Using Deep Learning Methods," in *2018 International Joint Conference on Neural Networks (IJCNN), 2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro, Brazil, July 2018), 1–8, https://doi.org/10.1109/IJCNN.2018.8489369.

[33] T. P. Moreira, M. L. Perez, R. d O. Werneck, and E. Valle, "Where Is My Puppy? Retrieving Lost Dogs by Facial Features," *Multimedia Tools and Applications* 76, no. 14 (2017): 15325–15340, https://doi.org/10.1007/s11042-016-3824-1.

[34] C. Bergler, H. Schröter, R. X. Cheng, et al., "ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning," *Scientific Reports* 9, no. 1 (2019): 10997–11017, https://doi.org/10.1038/s41598-019-47335-w.

[35] A. Nasirahmadi, J. Gonzalez, B. Sturm, O. Hensel, and U. Knierim, "Pecking Activity Detection in Group-Housed Turkeys Using Acoustic Data and a Deep Learning Technique," *Biosystems Engineering* 194 (2020): 40–48, https://doi.org/10.1016/j.biosystemseng.2020.03.015.

[36] Z. Cao, S. Yu, B. Ouyang, et al., "Marine Animal Classification With Correntropy-Loss-Based Multiview Learning," *IEEE Journal of Oceanic Engineering* 44, no. 4 (2019): 1116–1129, https://doi.org/10.1109/joe.2018.2861500.

[37] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, et al., "Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images With Deep Learning," *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 25 (2018): E5716–E5725, https://doi.org/10.1073/pnas.1719367115.

[38] R. Chen, R. Little, L. Mihaylova, R. Delahay, and R. Cox, "Wildlife Surveillance Using Deep Learning Methods," *Ecology and Evolution* 9, no. 17 (2019): 9453–9466, https://doi.org/10.1002/ece3.5410.

[39] H. Nguyen, S. J. Maclagan, T. D. Nguyen, et al., "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)* (Tokyo, Japan, October 2017), 40–49, https://doi.org/10.1109/DSAA.2017.31.

[40] D. Schofield, A. Nagrani, A. Zisserman, et al., "Chimpanzee Face Recognition From Videos in the Wild Using Deep Learning," *Science Advances* 5, no. 9 (2019): eaaw0736, https://doi.org/10.1126/sciadv.aaw0736.

[41] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, and Y. Yagi, "RGB-D Video-Based Individual Identification of Dairy Cows Using Gait and Texture Analyses," *Computers and Electronics in Agriculture* 165 (2019): 104944, https://doi.org/10.1016/j.compag.2019.104944.

[42] A. Delplanque, S. Foucher, P. Lejeune, J. Linchant, and J. Théau, "Multispecies Detection and Identification of African Mammals in Aerial Imagery Using Convolutional Neural Networks," *Remote Sensing in Ecology and Conservation* 8, no. 2 (2021): 166–179, https://doi.org/10.1002/rse2.234.

[43] E. T. Psota, M. Mittek, L. C. Pérez, T. Schmidt, and B. Mote, "Multi-Pig Part Detection and Association With a Fully-Convolutional Network," *Sensors* 19, no. 4 (2019): 852, https://doi.org/10.3390/s19040852.

[44] H. S. Demir, J. B. Christen, and S. Ozev, "Energy-Efficient Image Recognition System for Marine Life," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39, no. 11 (2020): 3458–3466, https://doi.org/10.1109/TCAD.2020.3012745.

[45] N. Ramesh and T. Tasdizen, "Cell Segmentation Using a Similarity Interface With a Multi-Task Convolutional Neural Network," *IEEE journal of biomedical and health informatics* 23, no. 4 (2019): 1457–1468, https://doi.org/10.1109/JBHI.2018.2885544.

[46] E. M. Ditria, S. Lopez-Marcano, M. Sievers, E. L. Jinks, C. J. Brown, and R. M. Connolly, "Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning," *Frontiers in Marine Science* 7 (2020): 429, https://doi.org/10.3389/fmars.2020.00429.

[47] H. T. Rauf, M. I. U. Lali, S. Zahoor, S. Z. H. Shah, A. U. Rehman, and S. A. C. Bukhari, "Visual Features Based Automated Identification of Fish Species Using Deep Convolutional Neural Networks," *Computers and Electronics in Agriculture* 167 (2019): 105075, https://doi.org/10.1016/j.compag.2019.105075.

[48] S. Meena and L. Agilandeeswari, "An Efficient Framework for Animal Breeds Classification Using Semi-Supervised Learning and Multi-Part Convolutional Neural Network (MP-CNN)," *IEEE Access* 7 (2019): 151783–151802, https://doi.org/10.1109/ACCESS.2019.2947717.

[49] J. Rodríguez Alvarez, M. Arroqui, P. Mangudo, et al., "Body Condition Estimation on Cows From Depth Images Using Convolutional Neural Networks," *Computers and Electronics in Agriculture* 155 (2018): 12–22, https://doi.org/10.1016/j.compag.2018.09.039.

[50] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, "Workflow and Convolutional Neural Network for Automated Identification of Animal Sounds," *Ecological Indicators* 124 (2021): 107419, https://doi.org/10.1016/j.ecolind.2021.107419.

[51] N. K. El Abbadi and E. M. T. A. Alsaadi, "An Automated Vertebrate Animals Classification Using Deep Convolution Neural Networks," in *2020 International Conference on Computer Science and Software Engineering (CSASE)* (Duhok, Iraq: IEEE, April 2020), 72–77, https://doi.org/10.1109/CSASE48920.2020.9142070.

[52] K. Saleh, M. Hossny, and S. Nahavandi, "Effective Vehicle-Based Kangaroo Detection for Collision Warning Systems Using Region-Based Convolutional Networks," *Sensors* 18, no. 6 (2018): 1913, https://doi.org/10.3390/s18061913.

[53] W. Fang, Y. Ding, F. Zhang, and V. S. Sheng, "DOG: A New Background Removal for Object Recognition from Images," *Neurocomputing* 361 (2019): 85–91, https://doi.org/10.1016/j.neucom.2019.05.095.

[54] S. Schneider, S. Greenberg, G. W. Taylor, and S. C. Kremer, "Three Critical Factors Affecting Automated Image Species Recognition Performance for Camera Traps," *Ecology and*

Evolution 10, no. 7 (2020): 3503–3517, https://doi.org/10.1002/ece3.6147.

[55] A. C. Ferreira, L. R. Silva, F. Renna, et al., "Deep Learning-Based Methods for Individual Recognition in Small Birds," *Methods in Ecology and Evolution* 11, no. 9 (2020): 1072–1085, https://doi.org/10.1111/2041-210x.13436.

[56] S. T. Schwartz and M. E. Alfaro, "Sashimi: A Toolkit for Facilitating High-Throughput Organismal Image Segmentation Using Deep Learning," *Methods in Ecology and Evolution* 12, no. 12 (2021): 2341–2354, https://doi.org/10.1111/2041-210X.13712.

[57] P. C. Gray, A. B. Fleishman, D. J. Klein, et al., "A Convolutional Neural Network for Detecting Sea Turtles in Drone Imagery," *Methods in Ecology and Evolution* 10, no. 3 (2019): 345–355, https://doi.org/10.1111/2041-210X.13132.

[58] S. Shahinfar, P. D. Meek, and G. Falzon, "How Many Images Do I Need? Understanding How Sample Size Per Class Affects Deep Learning Model Performance Metrics for Balanced Designs in Autonomous Wildlife Monitoring," *Ecological Informatics* 57 (2020): 101085, https://doi.org/10.1016/j.ecoinf.2020.101085.

[59] Y. Guo, T. A. Rothfus, A. S. Ashour, L. Si, C. Du, and T. F. Ting, "Varied Channels Region Proposal and Classification Network for Wildlife Image Classification Under Complex Environment," *IET Image Processing* 14, no. 4 (2020): 585–591, https://doi.org/10.1049/iet-ipr.2019.1042.

[60] A. Shepley, G. Falzon, P. Meek, and P. Kwan, "Automated Location Invariant Animal Detection in Camera Trap Images Using Publicly Available Data Sources," *Ecology and Evolution* 11, no. 9 (2021): 4494–4506, https://doi.org/10.1002/ece3.7344.

[61] A. Guzman-Pando and M. I. Chacon-Murguia, "DeepFoveaNet: Deep Fovea Eagle-Eye Bioinspired Model to Detect Moving Objects," *IEEE Transactions on Image Processing* 30 (2021): 7090–7100, https://doi.org/10.1109/TIP.2021.3101398.

[62] R. Singh and V. Mumbarekar, "Neural Network Model Approach for Automated Benthic Animal Identification," *ICT Express* 8, no. 4 (2022): 640–645, https://doi.org/10.1016/j.icte.2021.03.003.

[63] L. Han, P. Tao, and R. R. Martin, "Livestock Detection in Aerial Images Using a Fully Convolutional Network," *Computational Visual Media* 5, no. 2 (2019): 221–228, https://doi.org/10.1007/s41095-019-0132-5.

[64] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, et al., "Machine Learning to Classify Animal Species in Camera Trap Images: Applications in Ecology," *Methods in Ecology and Evolution* 10, no. 4 (2019): 585–590, https://doi.org/10.1111/2041-210X.13120.

[65] X. Han and R. Jin, "A Small Sample Image Recognition Method Based on ResNet and Transfer Learning," in *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)* (Beijing, China: IEEE, June 2020), 76–81, https://doi.org/10.1109/ICCIA49625.2020.00022.

[66] S. Taheri and Ö Toygar, "Animal Classification Using Facial Images With Score-Level Fusion," *IET Computer Vision* 12, no. 5 (2018): 679–685, https://doi.org/10.1049/iet-cvi.2017.0079.

[67] S. J. Hong, Y. Han, S. Y. Kim, A. Y. Lee, and G. Kim, "Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery," *Sensors* 19, no. 7 (2019): 1651, https://doi.org/10.3390/s19071651.

[68] A. Tan, G. Zhou, and M. He, "Rapid Fine-Grained Classification of Butterflies Based on FCM-KM and Mask R-CNN Fusion," *IEEE Access* 8 (2020): 124722–124733, https://doi.org/10.1109/ACCESS.2020.3007745.

[69] T. T. Høye, J. Ärje, K. Bjerge, et al., "Deep Learning and Computer Vision Will Transform Entomology," *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 2 (2021): e2002545117, https://doi.org/10.1073/pnas.2002545117.

[70] S. Thangavel and C. S. Shokkalingam, "The IoT Based Embedded System for the Detection and Discrimination of Animals to Avoid Human–Wildlife Conflict," *Journal of Ambient Intelligence and Humanized Computing* 13, no. 6 (2022): 3065–3081, https://doi.org/10.1007/s12652-021-03141-9.

[71] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a Percent of Labels Is Enough: Efficient Animal Detection in UAV Imagery Using Deep CNNs and Active Learning," *IEEE Transactions on Geoscience and Remote Sensing* 57, no. 12 (2019): 9524–9533, https://doi.org/10.1109/tgrs.2019.2927393.

[72] A. Ulhaq, P. Adams, T. E. Cox, A. Khan, T. Low, and M. Paul, "Automated Detection of Animals in Low-Resolution Airborne Thermal Imagery," *Remote Sensing* 13, no. 16 (2021): 3276, https://doi.org/10.3390/rs13163276.

[73] B. Kellenberger, D. Marcos, and D. Tuia, "Detecting Mammals in UAV Images: Best Practices to Address a Substantially Imbalanced Dataset With Deep Learning," *Remote Sensing of Environment* 216 (2018): 139–153, https://doi.org/10.1016/j.rse.2018.06.028.

[74] W. Van Eeden, J. P. De Villiers, R. Berndt, W. A. Nel, and E. Blasch, "Micro-Doppler Radar Classification of Humans and Animals in an Operational Environment," *Expert Systems With Applications* 102 (2018): 1–11, https://doi.org/10.1016/j.eswa.2018.02.019.

[75] S. Kumar and S. K. Singh, "Monitoring of Pet Animal in Smart Cities Using Animal Biometrics," *Future Generation Computer Systems* 83 (2018): 553–563, https://doi.org/10.1016/j.future.2016.12.006.

[76] V. Lopez-Vazquez, J. M. Lopez-Guede, D. Chatzievangelou, and J. Aguzzi, "Deep Learning Based Deep-Sea Automatic Image Enhancement and Animal Species Classification," *Journal of Big Data* 10, no. 1 (2023): 37–26, https://doi.org/10.1186/s40537-023-00711-w.

[77] M. Willi, R. T. Pitman, A. W. Cardoso, et al., "Identifying Animal Species in Camera Trap Images Using Deep Learning and Citizen Science," *Methods in Ecology and Evolution* 10, no. 1 (2019): 80–91, https://doi.org/10.1111/2041-210X.13099.

[78] V. Miele, G. Dussert, B. Spataro, S. Chamaillé-Jammes, D. Allainé, and C. Bonenfant, "Revisiting Animal Photo-Identification Using Deep Metric Learning and Network Analysis," *Methods in Ecology and Evolution* 12, no. 5 (2021): 863–873, https://doi.org/10.1111/2041-210X.13577.

[79] W. Xu, X. Zhang, L. Yao, W. Xue, and B. Wei, "A Multi-View CNN-Based Acoustic Classification System for Automatic Animal Species Identification," *Ad Hoc Networks* 102 (2020): 102115, https://doi.org/10.1016/j.adhoc.2020.102115.

[80] H. Yousif, J. Yuan, R. Kays, and Z. He, "Object Detection from Dynamic Scene Using Joint Background Modeling and Fast Deep Learning Classification," *Journal of Visual Communication and Image Representation* 55 (2018): 802–815, https://doi.org/10.1016/j.jvcir.2018.08.013.

[81] Z. Zhao, Y. Liu, X. Sun, J. Liu, X. Yang, and C. Zhou, "Composited FishNet: Fish Detection and Species Recognition From Low-Quality Underwater Videos," *IEEE*

Transactions on Image Processing 30 (2021): 4719–4734, https://doi.org/10.1109/TIP.2021.3074738.

[82] S. Seibold, M. M. Gossner, N. K. Simons, et al., "Arthropod Decline in Grasslands and Forests Is Associated With Landscape-Level Drivers," *Nature* 574, no. 7780 (2019): 671–674, https://doi.org/10.1038/s41586-019-1684-3.

[83] S. Ness, *The Orchive: A System for Semi-Automatic Annotation and Analysis of a Large Collection of Bioacoustic Recordings* (Victoria, Canada: University of Victoria, 2013).

[84] H. Hu, Z. Sun, and L. Su, "Underwater Motion and Activity Recognition Using Acoustic Wireless Networks," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (Dublin, Ireland: IEEE, June 2020), 1–7, https://doi.org/10.1109/ICC40277.2020.9148998.

[85] H. Strager, "Pod-Specific Call Repertoires and Compound Calls of Killer Whales, *Orcinus orca* Linnaeus, 1758, in the Waters of Northern Norway," *Canadian Journal of Zoology* 73, no. 6 (1995): 1037–1047, https://doi.org/10.1139/z95-124.

[86] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer, "Snapshot Serengeti, High-Frequency Annotated Camera Trap Images of 40 Mammalian Species in an African Savanna," *Scientific Data* 2, no. 1 (2015): 150026–150114, https://doi.org/10.1038/sdata.2015.26.

[87] A. Lopes, R. Souza, and H. Pedrini, "A Survey on RGB-D Datasets," *Computer Vision and Image Understanding* 222 (2022): 103489, https://doi.org/10.1016/j.cviu.2022.103489.

[88] N. Rey, M. Volpi, S. Joost, and D. Tuia, "Detecting Animals in African Savanna With UAVs and the Crowds," *Remote Sensing of Environment* 200 (2017): 341–351, https://doi.org/10.1016/j.rse.2017.08.026.

[89] G. M. Schuster and A. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding* (Berlin, Germany: Springer Science & Business Media, 2013).

[90] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication* (Cambridge, UK: Cambridge University Press, 2005).

[91] M. I. Jordan and T. M. Mitchell, "Machine Learning: Trends, Perspectives, and Prospects," *Science* 349, no. 6245 (2015): 255–260, https://doi.org/10.1126/science.aaa8415.

[92] S. Abney, *Semisupervised Learning for Computational Linguistics* (Boca Raton, FL: Chapman and Hall/CRC, 2007).

[93] T. Y. Lin, M. Maire, S. Belongie, et al., "Microsoft Coco: Common Objects in Context," in *European conference on computer vision* (Berlin, Germany: Springer, May 2014), 740–755, https://doi.org/10.48550/arXiv.1405.0312.

[94] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: A Novel Image Dataset for Benchmarking Machine Learning Algorithms," (2017), https://arxiv.org/abs/1708.07747, https://doi.org/10.48550/arXiv.1708.07747.

[95] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A Large-Scale Hierarchical Image Database," in *2009 IEEE conference on computer vision and pattern recognition IEEE* (Miami, FL, June 2009), 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

[96] Q. Zhang, K. Ahmed, N. Sharda, and H. Wang, "Australian Animal Species Selection and Image Data Collection," in *2023 27th International Conference Information Visualisation (IV)* (Tampere, Finland: IEEE, July 2023), 55–63, https://doi.org/10.1109/IV60283.2023.00020.

[97] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI, July 2017), 936–944, https://doi.org/10.1109/CVPR.2017.106.

[98] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification With Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems* 25 (2012).

[99] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," https://arxiv.org/abs/1409.1556, https://doi.org/10.48550/arXiv.1409.1556.

[100] O. Russakovsky, J. Deng, H. Su, et al., "Imagenet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision* 115, no. 3 (2015): 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[101] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Berlin, Germany: Springer, 2006).

[102] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," *Artificial Intelligence* 90 (2002).

[103] L. Silveira, A. T. A. Jácomo, and J. A. F. Diniz-Filho, "Camera Trap, Line Transect Census and Track Surveys: A Comparative Evaluation," *Biological Conservation* 114, no. 3 (2003): 351–355, https://doi.org/10.1016/S0006-3207(03)00063-6.

[104] C. Mora, D. P. Tittensor, S. Adl, A. G. Simpson, and B. Worm, "How Many Species Are There on Earth and in the Ocean?" *PLoS Biology* 9, no. 8 (2011): e1001127, https://doi.org/10.1371/journal.pbio.1001127.

[105] E. Australia, *Australian Biological Resources Study* (Canberra, Australia: Environment Australia, 1998).

[106] D. Orlovich, "iNaturalist," (2024), https://www.gbif.org/occurrence/4516483094.

[107] C. Linnaeus, "Systema Naturae," *Stockholm Laurentii Salvii* 1 (1758).

[108] S. Zheng, C. Zhou, X. Jiang, J. Huang, and D. Xu, "Progress on Infrared Imaging Technology in Animal Production: A Review," *Sensors* 22, no. 3 (2022): 705, https://doi.org/10.3390/s22030705.

[109] S. H. Wu, H. W. Chang, R. S. Lin, and M. N. Tuanmu, "SILIC: A Cross Database Framework for Automatically Extracting Robust Biodiversity Information From Soundscape Recordings Based on Object Detection and a Tiny Training Dataset," *Ecological Informatics* 68 (2022): 101534, https://doi.org/10.1016/j.ecoinf.2021.101534.

[110] J. Xiao, G. Liu, K. Wang, and Y. Si, "Cow Identification in Free-Stall Barns Based on an Improved Mask R-CNN and an SVM," *Computers and Electronics in Agriculture* 194 (2022): 106738, https://doi.org/10.1016/j.compag.2022.106738.

[111] J. Lu, P. Gong, J. Ye, and C. Zhang, "Learning From Very Few Samples: A Survey," https://arxiv.org/abs/2007.15484, https://doi.org/10.48550/arXiv.2007.15484.

[112] J. Webster, *Animal Welfare: Limping Towards Eden: A Practical Approach to Redressing the Problem of Our Dominion Over the Animals* (Hoboken, NJ: John Wiley & Sons, 2008).