

Article

# Macao-ebird: A Curated Dataset for Artificial-Intelligence-Powered Bird Surveillance and Conservation in Macao

Xiaoyuan Huang<sup>1,2</sup> , Silvia Mirri<sup>3</sup>  and Su-Kit Tang<sup>1,\*</sup> 

<sup>1</sup> Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR 999078, China; huangxy@dgut.edu.cn

<sup>2</sup> School of Electrical Engineering and Intelligentization, Dongguan University of Technology, Dongguan 523000, China

<sup>3</sup> Department of Computer Science and Engineering, University of Bologna, 40128 Bologna, Italy; silvia.mirri@unibo.it

\* Correspondence: sktang@mpu.edu.mo

**Abstract:** Artificial intelligence (AI) currently exhibits considerable potential within the realm of biodiversity conservation. However, high-quality regionally customized datasets remain scarce, particularly within urban environments. The existing large-scale bird image datasets often lack a dedicated focus on endangered species endemic to specific geographic regions, as well as a nuanced consideration of the complex interplay between urban and natural environmental contexts. Therefore, this paper introduces *Macao-ebird*, a novel dataset designed to advance AI-driven recognition and conservation of endangered bird species in Macao. The dataset comprises two subsets: (1) *Macao-ebird-cls*, a classification dataset with 7341 images covering 24 bird species, emphasizing endangered and vulnerable species native to Macao; and (2) *Macao-ebird-det*, an object detection dataset generated through AI-agent-assisted labeling using grounding DETR with improved denoising anchor boxes (DINO), significantly reducing manual annotation effort while maintaining high-quality bounding-box annotations. We validate the dataset's utility through baseline experiments with the You Only Look Once (YOLO) v8–v12 series, achieving a mean average precision (mAP50) of up to 0.984. *Macao-ebird* addresses critical gaps in the existing datasets by focusing on region-specific endangered species and complex urban–natural environments, providing a benchmark for AI applications in avian conservation.



Academic Editor: Rüdiger Pryss

Received: 18 April 2025

Revised: 17 May 2025

Accepted: 27 May 2025

Published: 30 May 2025

**Citation:** Huang, X.; Mirri, S.; Tang, S.-K. *Macao-ebird: A Curated Dataset for Artificial-Intelligence-Powered Bird Surveillance and Conservation in Macao*. *Data* **2025**, *10*, 84. <https://doi.org/10.3390/data10060084>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** endangered bird; *Macao-ebird* dataset; classification; detection; grounding DINO; YOLO

## 1. Introduction

Earth's biodiversity is facing an unprecedented crisis, with species extinction rates accelerating due to habitat loss, climate change, pollution, and overexploitation [1,2]. Birds, as key indicators of environmental health, are particularly vulnerable to these threats, with many species experiencing significant population declines and range contractions [3,4]. Understanding and monitoring bird species and populations are crucial for effective conservation strategies, enabling targeted interventions to protect vulnerable species and their habitats [5]. The traditional methods of bird identification and monitoring often rely on expert ornithologists, which can be time-consuming, costly, and limited in geographic scope [6]. Citizen science initiatives, where volunteers contribute data on bird sightings, have become increasingly valuable, but they can be subject to biases and require rigorous quality control [7,8].

Artificial intelligence (AI), particularly computer vision techniques, offer a promising alternative for automated bird identification and monitoring [9,10]. Deep learning (DL) models, such as convolutional neural networks (CNNs) and object detection algorithms, have demonstrated remarkable accuracy in identifying bird species from images and videos [11,12]. However, the success of these AI-based methods critically depends on the availability of large, high-quality labeled datasets [13]. The existing bird datasets, such as *Caltech-UCSD Birds 200 (CUB-200)* [14] and the *iNaturalist* dataset [15], are valuable resources. Nevertheless, they may not adequately represent the specific bird species found in particular geographic regions, especially those with unique ecological characteristics. Furthermore, these datasets often lack sufficient data on endangered species, which are a priority for conservation efforts.

Macao, despite its small size and high population density, is an important stopover point for migratory birds and supports a surprisingly diverse range of avian species [16,17]. The unique blend of urban and natural environments in Macao creates both opportunities and challenges for bird conservation. However, information about Macao's avifauna is scattered, not easily accessible, and often lacks the detail needed for effective conservation planning. In addition, the urban landscape presents a significant threat, impacting habitats and presenting obstacles for migrating species [18].

To address these challenges, we introduce *Macao-ebird*, a novel dataset specifically designed for bird species identification in Macao, with a particular focus on endangered species. *Macao-ebird* aims to provide a valuable resource for researchers and practitioners interested in developing AI-powered tools for avian conservation in this unique urban environment. This paper makes the following contributions:

1. *Macao-ebird-cl*s: A classification dataset containing 7341 images of 24 bird species by data crawling, emphasizing endangered and vulnerable species in Macao. This dataset is specifically curated to reflect the local avifauna and conservation priorities.
2. *Macao-ebird-det*: An object detection dataset created using AI-agent-assisted labeling via grounding DINO (DETR with improved denoising anchor boxes) [19], providing bounding-box annotations for bird instances. This innovative approach significantly reduces the manual effort required for creating a large-scale detection dataset.
3. Baseline Experiments: A demonstration of *Macao-ebird-det*'s utility through baseline experiments using You Only Look Once (YOLO) v8–v12 [20–24] for bird species detection. These experiments provide a benchmark for future research and development using the dataset.

## 2. Related Works

### 2.1. Bird Datasets

Publicly available datasets have been instrumental in advancing avian research. The *CUB-200-2011* dataset [14], containing 11,788 images of 200 North American bird species, became a benchmark for fine-grained classification. Similarly, *iNaturalist 2017* [25] and *BirdCLEF* [26] extended coverage to global species with audio–visual data. However, these datasets lack annotations for endangered birds in specific geographic regions. For instance, Macao's critically endangered Black-faced Spoonbill (*Platalea minor*) is under-represented in the existing resources. Recent efforts like the *iNaturalist 2021 Birds Dataset* [27] partially address this gap by crowdsourcing images, but domain-specific challenges (e.g., small sample sizes and occlusions) remain unresolved. To the best of our knowledge, there is no bird dataset specifically for Macao, especially for endangered or nationally protected birds.

## 2.2. AI-Agent-Assisted Labeling

The existing assisted labeling methods, such as active learning [28] and semi-supervised techniques [29], reduce manual effort but remain constrained by predefined taxonomies, failing to detect novel or rare species (open-vocabulary detection). Interactive tools like the segment anything model (SAM) [30] require heavy human input for fine-grained species differentiation while lacking semantic alignment between visual features and ecological knowledge. Grounding DINO addresses these gaps through text-guided open-set detection. By aligning image regions with text embeddings (e.g., “find the bird *Platalea minor*.”), it enables zero-shot localization of unseen species without bounding-box annotations, significantly reducing dependency on labeled data—critical for endangered avian conservation. By leveraging text-guided localization, it reduces manual bounding-box annotation efforts, cutting the labeling costs by 50–70% [31]. These advancements effectively address the core challenge of scarce annotated data for endangered avian monitoring and provide a scalable solution for biodiversity conservation.

## 2.3. Bird Species Classification and Detection

The traditional methods for bird species recognition relied on handcrafted features such as color histograms and texture descriptors (e.g., Gabor filters) [32] but struggled with lighting variations, cluttered backgrounds, and taxonomic similarities. DL then revolutionized the field: CNNs like ResNet [33] and EfficientNet [34] achieved 85–98% accuracy on large datasets by learning hierarchical visual patterns, while YOLO-based architectures [35–38] enabled real-time detection of small objects [39,40].

Recent advances addressed data scarcity through transfer learning. Vision transformers (ViTs) like data-efficient image transformers (DeiT) [41] and pre-trained models like BirdNET [42] leveraged domain knowledge from ImageNet to improve low-data performance. Data augmentation techniques such as CutMix [43], MixUp [44], and generative adversarial networks (GANs) generated images [45] that partially mitigated the data limitations but required careful annotation. Multimodal approaches combined acoustic and visual data: BirdCLEF challenges [46] integrated sound for urban bird recognition, while studies demonstrated the potential of audio–visual sensors to detect endangered species like *Platalea minor* [47,48]. However, while global datasets like Birdsnap [49] and CUB-200-2011 provide broad taxonomic coverage, there is a lack of Macao-specific datasets of endemic or protected species for training the DL models. Less than 5% of the existing studies focus on South China’s avian fauna, with none tailored to Macao’s unique urban ecosystems [50]. This data scarcity not only limits the model accuracy in urbanized regions like Macao but also hinders practical applications for critical conservation tasks—such as monitoring habitat fragmentation effects on migratory birds [51].

## 3. Dataset Creation

The *Macao-ebird* dataset was constructed using a combination of publicly available online resources and the AI-agent-assisted labeling process. We prioritized species listed as endangered or vulnerable by *The Catalogue of Birds in the Cotai Ecological Zone* [52] and *Report on the State of Macao’s Environment* [53], which is published from Macao SAR Government Environmental Protection Bureau. These species belong to *China Red Data Book of Endangered Animals* [54] or *China Protection Animal Classes I and II* [55]. This selection aimed to create a dataset focused on species most in need of monitoring and conservation efforts.

### 3.1. *Macao-ebird-cl*s

#### 3.1.1. Image Acquisition via Web Scraping

Ornithology and bird conservation efforts are increasingly reliant on digital resources that facilitate data collection, identification, and comprehensive species understanding. Images for the *Macao-ebird-cl*s dataset were initially acquired using web scraping methodologies from two primary sources. We leveraged global bird observation and recording platforms, specifically eBird (Cornell Lab of Ornithology) and Observation.org. Images from these platforms generally came with initial species labels assigned by the users who submitted the observation records. We also utilized Baidu Image Search. For this source, candidate images were retrieved by executing search queries based on the scientific or common names of the target bird species. The initial presumptive labels for these images were thus inferred from the search terms or accompanying image metadata and captions.

For citizen science platforms such as eBird and Observation.org, data (including images) are typically contributed by users under specific terms that often permit use for research and non-commercial purposes, frequently adhering to Creative Commons licenses. We have complied with the terms of use of these platforms. For images obtained through general web searches (e.g., Baidu Image Search), the copyright for these images remains with the original creators. Our dataset has been compiled for non-commercial research and educational purposes, with the aim of advancing AI-based bird monitoring and conservation, which also aligns with the spirit of data sharing within the scientific and conservation communities.

#### 3.1.2. Image Curation and Selection

After scraping the images, we implemented image curation and selection to ensure taxonomic accuracy. We initially removed duplicate images (based on file hash) and images below a certain resolution threshold (minimum dimensions of  $224 \times 224$  pixels). Human annotators then confirmed species identity, assessed image quality (clarity, lighting, and composition), removed occluded images (e.g., images where branches obscured the bird), and excluded images of captive birds. It is important to note that the dataset only retains images that contain only one instance (i.e., one image corresponds to one instance). The process reduces the number of new categories formed by bird combinations and avoids ambiguity caused by subsequent mixed-label annotations.

#### 3.1.3. Dataset Statistics

Table 1 shows the number of instances in the *Macao-ebird-cl*s dataset. We obtained 7341 images across 24 bird species. Overall, the number of images of most birds is between 200 and 400, showing that the order of magnitude of image records is relatively concentrated. The bird with the largest number is *Circus spilonotus* (412 instances), while the bird with the smallest number is *Halcyon smyrnensis* (340 instances).

**Table 1.** The instance numbers of *Macao-ebird-cl*s dataset.

ID	Bird Species	Number of Instances
0	<i>Accipiter nisus</i>	323
1	<i>Arenaria interpres</i>	251
2	<i>Calidris falcinellus</i>	260
3	<i>Calidris tenuirostris</i>	227
4	<i>Calliope calliope</i>	389
5	<i>Centropus sinensis</i>	281

**Table 1.** *Cont.*

ID	Bird Species	Number of Instances
6	<i>Circus spilonotus</i>	412
7	<i>Egetta eulophotes</i>	330
8	<i>Egretta sacra</i>	379
9	<i>Elanus caeruleus</i>	304
10	<i>Falco amurensis</i>	276
11	<i>Falco tinnunculus</i>	271
12	<i>Garrulax canorus</i>	280
13	<i>Halcyon smyrnensis</i>	340
14	<i>Hydrophasianus chirurgus</i>	251
15	<i>Leiothrix argenteauris</i>	358
16	<i>Leiothrix lutea</i>	386
17	<i>Limnodromus semipalmatus</i>	245
18	<i>Merops philippinus</i>	316
19	<i>Milvus migrans</i>	389
20	<i>Numenius arquata</i>	257
21	<i>Pandion haliaetus</i>	283
22	<i>Platalea leucorodia</i>	266
23	<i>Platalea minor</i>	267
	Total instances	7341

Figure 1 shows samples of 24 bird species of the *Macao-ebird-cls* dataset. The illustrations cover the life scenes of most endangered birds (shallow water, mudflats, branches, grass, sky, etc.). Their sizes and locations are also different, ensuring the diversity of samples.



**Figure 1.** Samples of 24 bird species of the *Macao-ebird-cls* dataset.

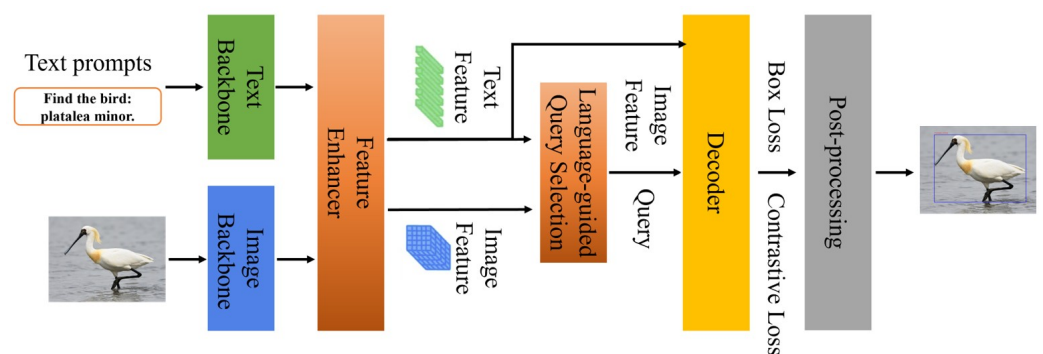
### 3.2. Macao-ebird-det

The images used for the detection dataset were the images collected for the classification dataset. We prioritized images that contained birds in complex environments to create a challenging and realistic detection task.

#### 3.2.1. AI-Agent-Assisted Labeling

Grounding DINO is an open-set object detector that bridges the gap between vision and language by enabling zero-shot object detection capabilities. Unlike traditional object detectors trained on fixed categories, grounding DINO can detect objects specified through natural language prompts, even those it has not seen during training. This is achieved by leveraging a transformer-based architecture trained on a large dataset of image–text pairs, enabling it to understand the semantic relationships between objects and their textual descriptions. It has become a popular choice for tasks requiring flexible object detection and semantic understanding of visual scenes.

Figure 2 is the multi-stage depiction of AI-agent-assisted labeling via grounding DINO. It begins with an input image and a natural language text prompt. The image is processed by a CNN-based image backbone (blue), extracting visual features, while the text prompt (e.g., “find the bird: *Accipiter nisus*”) is encoded by a transformer-based text backbone (green) into semantic embeddings. Subsequently, a feature enhancer (brown) refines both feature sets. Crucially, the language-guided query selection (brown) module employs cross-modal attention, enabling the text embeddings to guide the selection of relevant visual features, resulting in object queries. These queries are then fed into a decoder (yellow), which predicts object bounding boxes and confidence scores. Finally, a post-processing (gray) step refines the predictions via non-maximum suppression (NMS) and confidence thresholding, resulting in the output: the input image with overlaid bounding boxes accurately localizing the objects specified in the text prompt.



**Figure 2.** The multi-stage AI-agent-assisted labeling process via grounding DINO.

#### 3.2.2. Manual Verification and Refinement

The initial bounding-box annotations generated by grounding DINO were manually reviewed and refined by human annotators to ensure accuracy. This included correcting bounding-box positions, adding missing annotations (e.g., for birds that were partially obscured), and removing false positives (e.g., objects that were mistakenly identified as birds). We used a specialized annotation tool (e.g., LabelImg) to facilitate this process. We map the generated labels and bounding-box positions to the corresponding input images in batches and check them, including no labeled information or wrong labels and bounding-box positions.

The final instance counts for the *Macao-ebird-det* dataset are detailed in Table 2. From an initial set of 7341 input images encompassing 24 bird species (*Macao-ebird-cls* dataset), label and annotation generation was unsuccessful for 54 images. Consequently,

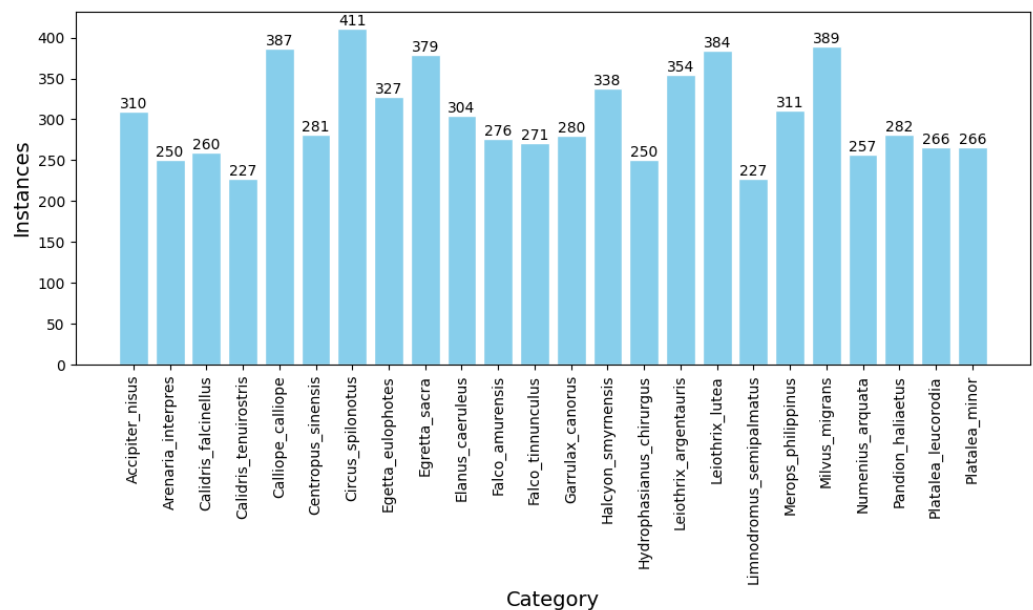
7287 images (instances) were successfully labeled, representing a labeling success rate of 99.18%. The methodology, which utilized bird folder categories from the classification dataset as input prompts, significantly reduced the likelihood of generating incorrect labels and bounding-box positions.

**Table 2.** The final instance counts for the *Macao-ebird-det* dataset by AI agent using grounding DINO.

Dataset	Number of Instances
<i>Macao-ebird-cls</i> (the original dataset)	7341
Labels generated failed	54
<i>Macao-ebird-det</i> (successfully labeled)	7287
Labeling success rate	99.18%

### 3.2.3. Dataset Statistics

Figure 3 includes the annotated instance numbers of *Macao-ebird-det* dataset with a total of 7287 images across 24 bird species. Image counts ranged from a minimum of 227 for *Limnodromus semipalmatus* and *Calidris tenuirostris* to a maximum of 411 for *Circus spilonotus*. The mean number of images per species was 307.96, with a median of 281.50. The dataset exhibits uniform class distribution with minimal outliers and negligible imbalance, demonstrating robust sampling methodology.



**Figure 3.** The annotated instance numbers of *Macao-ebird-det* dataset across 24 bird species.

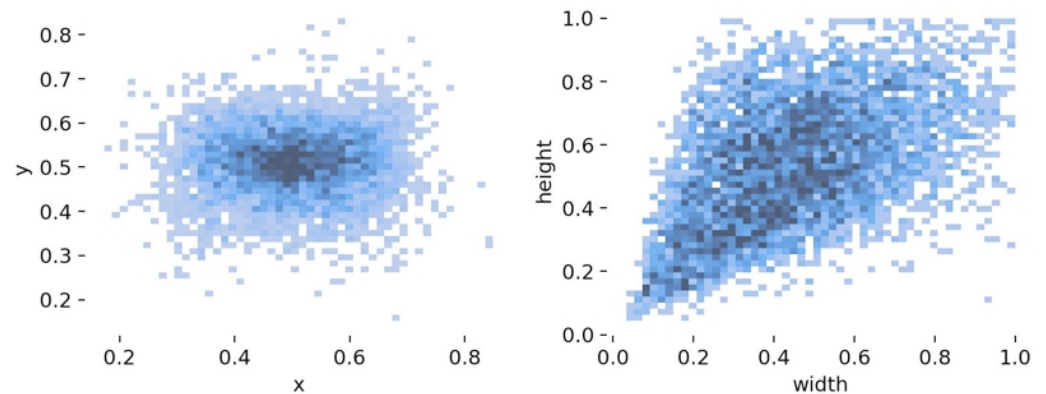
Figure 4 shows two 2D histograms (density scatter plots) visualizing characteristics of bounding-box distribution from *Macao-ebird-det* dataset.

The left plot shows the distribution of the normalized center coordinates ( $x, y$ ) of the bounding boxes. There is a strong concentration of data points (indicated by darker blue squares) around the center of the plot ( $x \approx 0.5, y \approx 0.5$ ). The density decreases radially outwards from this central point. This indicates that the objects detected in the dataset are predominantly located near the center of the image frames. Objects appearing near the edges of the images are less common.

This right plot shows the relationship between the normalized width and height of the bounding boxes. A clear positive correlation exists between width and height, visible as a diagonal trend extending from the bottom-left towards the top-right. The highest

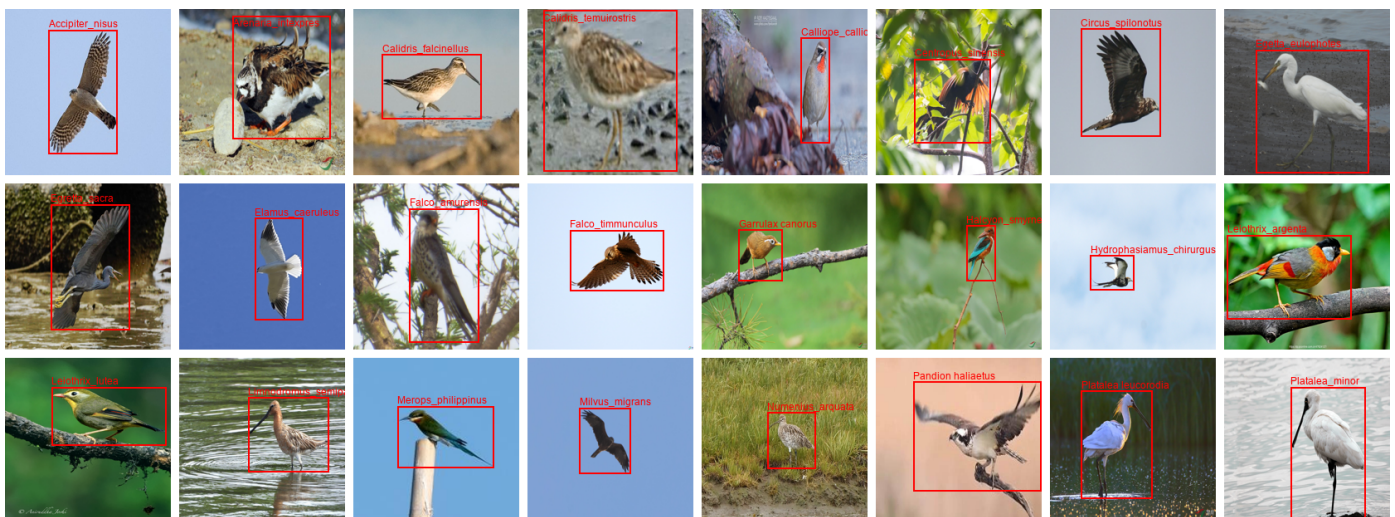
density of points are concentrated in the region where both width and height are relatively small (e.g., typically less than 0.4). The density decreases as either width or height (or both) increase. This signifies that width and height tend to scale together, which is expected as they represent the dimensions of physical objects. The plot strongly indicates that the dataset contains a large number of relatively small objects (small width and small height), while larger objects are less frequent. The correlation suggests a range of aspect ratios but generally avoids extreme cases like very wide but short boxes or very tall but thin boxes.

In summary, the visualizations suggest the *Macao-ebird-det* dataset primarily contains objects centered within the image frames and that most of these objects are relatively small in size, with width and height showing a positive correlation.



**Figure 4.** 2D histograms visualizing characteristics of bounding-box distribution from *Macao-ebird-det* dataset.

Figure 5 shows samples of annotated images of 24 bird species from the *Macao-ebird-det* dataset. As demonstrated, the birds in the samples are accurately detected. Bounding boxes effectively encompass their bodies and class labels correctly match the detected species. Even when birds exhibit diverse poses or are situated in complex environmental backgrounds, the AI agent still performs well in detection tasks.



**Figure 5.** Samples of annotated images of 24 bird species from the *Macao-ebird-det* dataset.

## 4. Baseline Experiments

### 4.1. Training Details

We utilized YOLOv8-v12 for our baseline experiments. Specifically, we selected the “n” and “s” versions of the models in each series for training and evaluation. These versions

have small computational complexity and low requirements on device computing power. They can be easily deployed to edge devices for real-time deployment and reasoning in the wild.

Baseline experiments are conducted in DL framework torch 1.12.1 with cu116 and trained on Graphics Processing Unit (GPU) NVIDIA GeForce 1080. The dataset The *Macao-ebird-det* dataset is divided into training, validation, and testing datasets in a 7:2:1 ratio. Specifically, for each bird species within the dataset, its corresponding images were randomly partitioned into 70% for the training dataset, 20% for the validation dataset, and 10% for the test dataset. The purpose of employing this stratified sampling strategy is to ensure class balance is maintained across the training, validation, and test datasets. The training phase utilizes transfer learning and data augmentation. Dataset images, after being loaded into memory during the training process, are all resized to 640x640 to ensure that the input image dimensions match those used during model training. The optimizer uses the auto mode with an initial learning rate of 0.01. dynamically adjusting the configuration during training. The batch size of a single GPU is 16, and the training round is 200 epochs. Giga floating point of operations (GFLOPs), precision, recall, and mean average precision (mAP50 and mAP50-90) are applied to evaluate the detection performance.

#### 4.2. Discussion

Table 3 includes the performance evaluation of the baseline experiments. We analyze and discuss the baseline experiments from four perspectives: annotation consistency, efficiency and inference speed, data distribution balance, and challenging case coverage.

**Table 3.** The performance evaluation of baseline experiments. All results are obtained using  $640 \times 640$  inputs.

Model	#Param. (M)	FLOPs (G)	Precision	Recall	mAP50	mAP50-95	Speed <sup>1</sup> (ms)
YOLOv8n	3.01	8.1	0.951	0.943	0.975	0.939	2.3
YOLOv8s	11.13	28.5	0.966	0.949	0.98	0.95	5.1
YOLOv9t	1.98	7.6	0.954	0.938	0.977	0.945	3.3
YOLOv9s	7.18	26.8	0.972	0.96	0.984	0.958	6.6
YOLOv10n	2.73	8.3	0.957	0.941	0.977	0.945	3.1
YOLOv10s	8.05	24.5	0.973	0.948	0.98	0.955	6.5
YOLOv11n	2.59	6.3	0.951	0.938	0.972	0.938	2.4
YOLOv11s	9.42	21.4	0.961	0.954	0.977	0.953	5.4
YOLOv12n	2.56	6.3	0.942	0.936	0.972	0.936	4
YOLOv12s	9.24	21.3	0.97	0.953	0.982	0.957	8.2

<sup>1</sup> Inference speed metrics refer to the model inference speed excluding pre-processing and post-processing. Speeds are averaged over *Macao-ebird-det* validation dataset using an NVIDIA GeForce GTX 1080 instance.

##### 4.2.1. Annotation Consistency Verification

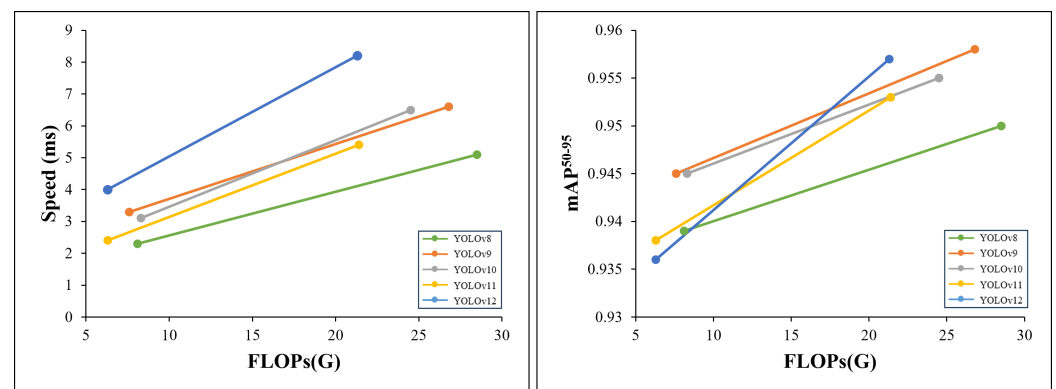
The lightweight models (e.g., YOLOv11n and YOLOv12n) exhibit stable mAP50 values (0.936–0.972), confirming minimal annotation noise. Cross-model consistency is further validated by all the models achieving  $mAP50 > 0.972$  (peaking at 0.984). This demonstrates precise bounding-box annotations with minimal localization errors at an Intersection over Union (IoU) threshold of 50%.

##### 4.2.2. Efficiency and Inference Speed Analysis

Figure 6 reveals the expected trade-offs between model size, computational cost (FLOPs), accuracy (mAP), and inference speed based on the corresponding results in Table 3. Within each version family (v8, v9, v10, v11, and v12), the ‘s’ models consistently

have more parameters and higher FLOPs than their ‘n’/‘t’ counterparts. This generally translates to higher accuracy (mAP50-95) but also significantly slower inference speeds (e.g., YOLOv8n: 2.3 ms vs. YOLOv8s: 5.1 ms; YOLOv11n: 2.4 ms vs. YOLOv11s: 5.4 ms). The ‘n’/‘t’ models offer rapid inference, with YOLOv8n (2.3 ms) and YOLOv11n (2.4 ms) being the fastest, suitable for resource-constrained environments. YOLOv12n (4 ms) is notably slower than the other nano models despite similar FLOPs to YOLOv11n.

While later series like v10/v11/v12 generally achieve high accuracy with competitive or lower FLOPs compared to earlier series (indicating good computational efficiency, e.g., YOLOv11s/12s have 21G FLOPs vs YOLOv8s’ 28.5G FLOPs for similar or better mAP50-95), their runtime efficiency (inference speed) does not show a monotonic improvement. For instance, YOLOv12s achieves the highest mAP50-95 (0.957) among the ‘s’ models shown but has the slowest inference time (8.2 ms). YOLOv11s offers a strong balance with high accuracy (0.953 mAP50-95) and relatively fast inference (5.4 ms) for an ‘s’ model.



**Figure 6.** Comparisons with different YOLO models in terms of their FLOP speed (**left**) and FLOP accuracy (**right**) trade-offs.

#### 4.2.3. Data Distribution Balance Verification

The dataset shows balanced class distribution, evidenced by precision–recall gaps  $< 0.02$  (e.g., YOLOv9s: 0.972 vs. 0.96) and consistent mAP50-95 improvements across model series (e.g., YOLOv8n→v8s: 0.939→0.95). It reflects diverse scenario coverage and effective distinction of the model capacities.

#### 4.2.4. Challenging Case Coverage Verification

A significant mAP50-95 variation (maximum difference is 0.022) highlights the dataset’s inclusion of challenging cases (such as small/dense, edge, or truncated bird samples), driving high-performance models (e.g., YOLOv9s) to adopt complex architectures for stability under stricter IoU thresholds. For YOLOv10s, precision (0.973) is significantly higher than recall (0.948). The possible reason is that the model filters out challenging samples more aggressively. Future work involves adding occluded/low-light samples to improve the distribution of challenging cases in the dataset.

In summary, the *Macao-ebird-det* dataset demonstrates excellent performance in annotation accuracy and covers a good range of scenarios, effectively supporting performance comparisons, particularly for lightweight models. The results highlight clear trade-offs between accuracy and speed. While the YOLOv9 series shows strong peak accuracy (YOLOv9s), and models like YOLOv11n/s demonstrate excellent runtime efficiency (inference speed), careful consideration of the specific balance between accuracy, computational cost (FLOPs), and real-world inference speed is necessary when selecting a model for a particular application. The data suggest that later-generation models improve computational efficiency (accuracy per FLOP), but runtime speed requires individual evaluation.

## 5. Applications

### 5.1. AI-Powered Bird Surveillance Recognition and Conservation

The *Macao-ebird* dataset's primary application lies in serving as foundational data for training and evaluating artificial intelligence models, such as convolutional neural networks (CNNs) and transformers, specifically for bird classification and detection tasks. The *Macao-ebird-cls* subset is designed for classification algorithm development, while the *Macao-ebird-det* subset is tailored for object detection, enabling the assessment of various detection algorithms like the YOLO series, Faster R-CNN, and DETR in accurately locating and identifying birds within real-world scenarios.

Given its focus on Macao's avifauna, particularly including endangered species, the dataset is especially valuable for developing automated bird identification systems targeted at Macao and ecologically similar regions like the Pearl River Delta. Deploying models trained on these data—for instance, on monitoring cameras, drones, or mobile applications—facilitates automatic real-time monitoring and identification of rare birds. This capability provides crucial technological support for conservation efforts, including population counting, habitat analysis, and threat alerting, thereby actively promoting the effective protection of endangered bird species.

### 5.2. Citizen Science and Environmental Education

Containing over 7300 images with species labels, the *Macao-ebird* dataset constitutes a rich visual library suitable for public science and environmental education. These images can be utilized to develop resources such as bird identification websites, mobile applications, and educational materials like brochures or exhibits. Such tools can effectively assist the general public, students, and birdwatching enthusiasts in learning to recognize the common and rare bird species found in Macao. The detailed and representative imagery of local bird species helps to deepen citizens' understanding of local biodiversity, particularly the status of bird populations and their habitats, consequently enhancing public awareness regarding the importance of ecological conservation and encouraging greater participation in related activities.

### 5.3. Algorithm Research and Benchmarking

*Macao-ebird* serves as a valuable resource for advancing algorithm research and establishing performance benchmarks, particularly in the domain of fine-grained visual categorization, as bird identification often requires distinguishing between visually similar species or subspecies. The dataset provides a solid foundation for testing, comparing, and evaluating the performance of various fine-grained image classification and recognition algorithms.

The baseline performance metrics reported in the paper offer a clear point of reference for future research. Researchers can leverage *Macao-ebird* to validate the efficacy of novel algorithms and compare different methodologies, thus fostering standardization and reproducibility within the field. Furthermore, the AI-assisted labeling approach employed (using grounding DINO) presents a case study for research into annotation efficiency and quality assessment.

## 6. Limitations

Although the *Macao-ebird* dataset prioritizes endangered and protected avian species (24 species) within Macao, it does not encompass the entirety of Macao's avifauna, as documented in sources like *The Catalogue of Birds in the Cotai Ecological Zone* [53,54] (which indicates up to 174 species). This restricted species coverage limits the dataset's direct applicability for comprehensive biodiversity assessments or the monitoring of all local bird

populations. Future research aiming for a holistic evaluation of Macao's avian ecology would necessitate dataset expansion to include a wider range of species.

Each image in the *Macao-ebird* dataset contains only one bird instance. This design choice simplifies image annotation and certain classification tasks. Real-world computer vision scenarios, especially in object detection and segmentation, frequently feature multiple object instances co-occurring within a single frame. Consequently, the dataset's utility is constrained for the direct development or evaluation of algorithms specifically designed to handle multi-object detection or counting involving multiple bird instances.

The image curation process favored clear, unobstructed, and adequately sized images, leading to the exclusion of significantly occluded samples. The dataset may have a limited proportion of hard examples, such as instances with severe occlusion, high target-background similarity (camouflage), or suboptimal lighting conditions. To enhance the generalization capability and robustness of models for real-world deployment, future iterations of the dataset should incorporate greater diversity regarding challenging samples (occluded/low-light samples) to improve the distribution of challenging cases.

## 7. Conclusions

*Macao-ebird* fills a critical gap by providing a dedicated dataset for endangered bird species recognition within the unique ecological context of Macao. The dataset comprises two distinct subsets: *Macao-ebird-cls* and *Macao-ebird-det*. *Macao-ebird-cls* contains 7341 images across 24 species, focusing on those considered to be protected or endangered within Macao. The object detection dataset *Macao-ebird-det* with bounding-box annotations was created from the same images using our AI-agent-assisted labeling approach via grounding DINO. This innovative labeling method, reducing manual effort while maintaining high annotation quality, offers a significant contribution to the field. Our baseline experiments using YOLOv8-v12 demonstrate the dataset's utility, achieving mAP50 scores of up to 0.984 and confirming minimal annotation noise. It is worth emphasizing that the dataset contains bird images of various species, scenes, and high resolutions, so it can be widely used in related research, such as bird identification, behavior analysis, and population assessment.

By bridging AI and conservation biology, *Macao-ebird* provides a scalable framework for protecting biodiversity in urbanized habitats. We are confident that *Macao-ebird* will facilitate the development of impactful solutions for protecting Macao's unique avian biodiversity. Future work will expand the dataset to include rare species, challenging cases (e.g., small objects and dense clusters), and multimodal data (e.g., audio recordings), enhancing its applicability in ecological monitoring.

**Author Contributions:** Conceptualization, X.H. and S.-K.T.; methodology, X.H.; software, X.H.; validation, X.H.; formal analysis, X.H.; investigation, X.H.; resources, X.H.; data curation, X.H.; writing—original draft preparation, X.H.; writing—review and editing, X.H. and S.-K.T.; visualization, X.H.; supervision, S.M. and S.-K.T.; funding acquisition, X.H. and S.-K.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the General Program for Social Development in Science and Technology of Dongguan, Grant No. 20221800901472; the Institutional Educational Quality Project of Dongguan University of Technology, Grant Nos. 202102072, 202302060, and 2024020105; and the Undergraduate Innovation and Entrepreneurship Training Program of Dongguan University of Technology, Grant Nos. 202311819010, S202311819065, S202311819066, and 202411819161.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The Macao-ebird dataset is available for download at <https://www.kaggle.com/datasets/huangxiaoyuan/macao-ebird/data>, accessed on 27 May 2025. Detailed instructions on how to use the dataset can be found at the provided link.

**Acknowledgments:** This work is supported in part by the research grant (RP/FCA-09/2023) offered by Macao Polytechnic University and in part by Faculty of Applied Sciences, Macao Polytechnic University (fca.e1b1.8606.2).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
DL	Deep learning
CNNs	Convolutional neural networks
CUB-200	Caltech-UCSD Birds
DINO	DETR with improved denoising anchor boxes
YOLO	You Only Look Once
SAM	Segment anything model
ViTs	Vision transformers
DeiT	Data-efficient image transformer
GAN	Generative adversarial network
NMS	Non-Maximum suppression
GFLOPs	Giga floating point of operations (GFLOPs)
mAP	mean Average Precision

## References

1. Jetz, W.; Thomas, G.H.; Joy, J.B.; Hartmann, K.; Mooers, A.O. The global diversity of birds in space and time. *Nature* **2012**, *491*, 444–448. [[CrossRef](#)] [[PubMed](#)]
2. Pimm, S.L.; Russell, G.J.; Gittleman, J.L.; Brooks, T.M. The future of biodiversity. *Science* **1995**, *269*, 347–350. [[CrossRef](#)] [[PubMed](#)]
3. Donald, P.F.; Sanderson, F.J.; Burfield, I.J.; Bierman, S.M.; Gregory, R.D.; Waliczky, Z. International conservation policy delivers benefits for birds in Europe. *Science* **2007**, *317*, 810–813. [[CrossRef](#)]
4. Luck, G.W.; Carter, A.; Smallbone, L. Changes in bird functional diversity across multiple land uses: Interpretations of functional redundancy depend on functional group identity. *PLoS ONE* **2013**, *8*, e63671. [[CrossRef](#)] [[PubMed](#)]
5. Sutherland, W.J. *Ecological Census Techniques: A Handbook*; Cambridge University Press: Cambridge, UK, 2006.
6. Huntington, H.P. Using traditional ecological knowledge in science: Methods and applications. *Ecol. Appl.* **2000**, *10*, 1270–1274. [[CrossRef](#)]
7. Sauermann, H.; Vohland, K.; Antoniou, V.; Balázs, B.; Göbel, C.; Karatzas, K.; Mooney, P.; Perelló, J.; Ponti, M.; Samson, R.; et al. Citizen science and sustainability transitions. *Res. Policy* **2020**, *49*, 103978. [[CrossRef](#)]
8. Bird, T.J.; Bates, A.E.; Lefcheck, J.S.; Hill, N.A.; Thomson, R.J.; Edgar, G.J.; Stuart-Smith, R.D.; Wotherspoon, S.; Krkosek, M.; Stuart-Smith, J.F.; et al. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **2014**, *173*, 144–154. [[CrossRef](#)]
9. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. **2017**, *60*, 84–90. [[CrossRef](#)]
11. Islam, S.; Khan, S.I.A.; Abedin, M.M.; Habibullah, K.M.; Das, A.K. Bird species classification from an image using VGG-16 network. In Proceedings of the 7th International Conference on Computer and Communications Management, Bangkok, Thailand, 27–29 July 2019; pp. 38–42.
12. Kumar, M.; Yadav, A.K.; Kumar, M.; Yadav, D. Bird species classification from images using deep learning. In Proceedings of the International Conference on Computer Vision and Image Processing, Nagpur, India, 4–6 November 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 388–401.
13. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

14. Wah, C.; Branson, S.; Welinder, P.; Perona, P.; Belongie, S. *The Caltech-UCSD Birds-200-2011 Dataset*; Technical Report; CNS-TR-2011-001; California Institute of Technology: Pasadena, CA, USA, 2011.
15. iNaturalist. California Academy of Sciences: Wildflower Watch. Available online: <https://www.inaturalist.org> (accessed on 26 July 2024).
16. Li, J.; Lin, F.; Yang, S.; Chen, Y. Spatial Planning Strategies for Urban Ecology and Heritage Conservation in Macau: An Investigation of Ultra-High-Density Cities. *Information* **2024**, *15*, 799. [[CrossRef](#)]
17. Ptak, R. The Avifauna of Macau: A Note on the Aomen jilüe. *Monum. Serica* **2009**, *57*, 195–230. [[CrossRef](#)]
18. Derrick, C. Heritage Protection, Tourism and Urban Planning in Macau. *China's Macau Transformed: Challenge and Development in the 21st Century*; City University of HK Press: Hong Kong, China, 2014; p. 297.
19. Liu, S.; Zeng, Z.; Ren, T.; Li, F.; Zhang, H.; Yang, J.; Jiang, Q.; Li, C.; Yang, J.; Su, H.; et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 38–55.
20. YOLOv8. Roboflow, Inc. Available online: <https://yolov8.com> (accessed on 26 February 2025).
21. Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. Yolov9: Learning what you want to learn using programmable gradient information. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–21.
22. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011.
23. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.
24. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524.
25. Van Horn, G.; Mac Aodha, O.; Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; Belongie, S. The inaturalist species classification and detection dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8769–8778.
26. Joly, A.; Botella, C.; Picek, L.; Kahl, S.; Goëau, H.; Deneu, B.; Marcos, D.; Estopinan, J.; Leblanc, C.; Larcher, T.; et al. Overview of lifeclef 2023: Evaluation of ai models for the identification and prediction of birds, plants, snakes and fungi. In Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages, Thessaloniki, Greece, 18–21 September 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 416–439.
27. iNaturalist. iNaturalist 2021 Competition Dataset, 2021. Available online: <https://www.kaggle.com/c/inaturalist-2021> (accessed on 4 March 2024).
28. Krishnakumar, A. *Active Learning Literature Survey*; Technical Report; University of California: Santa Cruz, CA, USA, 2007; Volume 42.
29. Sohn, K.; Berthelot, D.; Li, C.L.; Izmailov, P.; Goodfellow, I.; Bengio, Y. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In Proceedings of the Advances in Neural Information Processing Systems, virtual, 6–12 December 2020; Volume 33, pp. 596–608.
30. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–3 October 2023.
31. Simpson, R.; Page, K.R.; De Roure, D. Zooniverse: Observing the world's largest citizen science platform. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–11 April 2014; pp. 1049–1054.
32. Xie, J.; Zhu, M. Handcrafted features and late fusion with deep learning for bird sound classification. *Ecol. Inform.* **2019**, *52*, 74–81. [[CrossRef](#)]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
35. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
36. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
37. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [[CrossRef](#)]
38. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
39. Wu, J.; Xu, W.; He, J.; Lan, M. YOLO for penguin detection and counting based on remote sensing images. *Remote Sens.* **2023**, *15*, 2598. [[CrossRef](#)]
40. Farman, H.; Ahmed, S.; Imran, M.; Noureen, Z.; Ahmed, M. Deep learning based bird species identification and classification using images. *J. Comput. Biomed. Inform.* **2023**, *6*, 79–96.

41. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
42. Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* **2021**, *61*, 101236. [[CrossRef](#)]
43. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 6023–6032.
44. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
45. Alfatemi, A.; Jamal, S.A.; Paykari, N.; Rahouti, M.; Amin, R.; Chehri, A. Refining Bird Species Identification through GAN-Enhanced Data Augmentation and Deep Learning Models. *Procedia Comput. Sci.* **2024**, *246*, 548–557. [[CrossRef](#)]
46. Kahl, S.; Navine, A.; Denton, T.; Klinck, H.; Hart, P.; Glotin, H.; Goëau, H.; Vellinga, W.P.; Planqué, R.; Joly, A. Overview of BirdCLEF 2022: Endangered bird species recognition in soundscape recordings. In Proceedings of the Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, Bologna, Italy, 5–8 September 2022.
47. Lewis, T.C.; Vargas, I.G.; Williams, S.; Beckerman, A.P.; Childs, D.Z. Using passive acoustic monitoring to estimate the abundance of a critically endangered parrot, the great green macaw (*Ara ambiguus*). *bioRxiv* **2022**, bioRxiv:2022.12.29.519860.
48. Chen, Y.; Yu, Y.t.; Meng, F.; Deng, X.; Cao, L.; Fox, A.D. Migration routes, population status and important sites used by the globally threatened Black-faced Spoonbill (*Platalea minor*): A synthesis of surveys and tracking studies. *Avian Res.* **2021**, *12*, 1–17. [[CrossRef](#)]
49. Berg, T.; Liu, J.; Woo Lee, S.; Alexander, M.L.; Jacobs, D.W.; Belhumeur, P.N. Birdsnap: Large-scale fine-grained visual categorization of birds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2011–2018.
50. Lee, M.B.; Peabotuwage, I.; Gu, H.; Zhou, W.; Goodale, E. Factors affecting avian species richness and occupancy in a tropical city in southern China: Importance of human disturbance and open green space. *Basic Appl. Ecol.* **2019**, *39*, 48–56. [[CrossRef](#)]
51. Jiang, X.; Mao, D.; Zhen, J.; Wang, J.; Van de Voorde, T. Exploring the conservation of historic avian corridors under urbanization threats in China: A case study of egrets in the Greater Bay Area. *Sci. Total Environ.* **2024**, *948*, 174921. [[CrossRef](#)]
52. DSPA. *The Catalogue of Bird in the Cotai Ecological Zone*; Technical Report; Environmental Protection Bureau (DSPA), Macao SAR Government: Macao, China, 2018. Available online: <https://www.dspa.gov.mo/place3.aspx> (accessed on 3 March 2025).
53. DSPA. *Report on the State of Macao's Environment*; Report; Environmental Protection Bureau (DSPA), Macao SAR Government: Macao, China, 2023.
54. Guangmei, Z.; Qishan, W. *China Red List of Endangered Animals: Birds*; Science Press: Beijing, China, 1998.
55. Zhigang, J. China's key protected species lists, their criteria and management. *Biodivers. Sci.* **2019**, *27*, 698. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.