Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Detection and Re-Identification-Based Individual Identification Approach for Flamingo Population Counting

Kadir KAYA[1]

kkaya4214@gmail.com

Ahmet BAYRAM[2]
Hasan PAŞALI[3]

Research Article

Keywords: YOLOv8m, Re-Identification (Re-ID), High-Quality Segment Anything Model (HQ-SAM), Omni-Scale Network (OSNet)

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11620053/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Detection and Re-Identification-Based Individual Identification Approach for Flamingo Population Counting

Dr. Kadir KAYA[1] Ahmet BAYRAM[2] and Hasan PAŞALI[3]

[1] General Directorate of Nature Conservation and National Parks, Electrical and Electronics Engineer, 0009-0000-7865-415X

[2] General Directorate of Nature Conservation and National Parks, M.Sc. Forest Engineer

[3] General Directorate of Nature Conservation and National Parks, Veterinarian

Corresponding Author: kkaya4214@gmail.com

Abstract:

The unique identification of individual animals remains a critical research challenge in the technology-based detection and tracking of wildlife. Flamingos are among the most difficult species to distinguish at the individual level due to their similarities in shape, color, and pattern. In this study, a detection- and Re-Identification (Re-ID)-based individual identification approach was implemented to enable flamingo population counting based on unique individuals. The YOLO (You Only Look Once) method was employed for the real-time detection and tracking of flamingos in video streams. The detected flamingos were separated from the background using the High-Quality Segment Anything Model (HQ-SAM). Deep learning-based methods were then comparatively applied to the segmented flamingo images for Re-ID-based individual identification.

The highest detection performance was achieved by the YOLOv8m-based model at an altitude of 30 m over an area of approximately 1 ha. The YOLOv8m model achieved an mAP@0.50 of 96.4% and a processing speed of 14.51 FPS. For Re-ID-based individual identification, the best performance was achieved by the Omni-Scale Network (OSNet) using images acquired at an altitude of 30 m. OSNet achieved Rank-1, Rank-5, and mAP scores of 91.75%, 93.97%, and 72.49%, respectively. This study aims to contribute to wildlife inventory studies by preventing duplicate counts through Re-ID-based individual identification and enabling the estimation of population size based on the number of unique individuals.

Keywords: YOLOv8m, Re-Identification (Re-ID), High-Quality Segment Anything Model (HQ-SAM), Omni-Scale Network (OSNet).

I. INTRODUCTİON

With the rapidly changing climatic conditions, awareness of wildlife monitoring, detection, and conservation has been increasing steadily. Monitoring wildlife populations, closely observing endangered species, and promptly implementing conservation measures are of critical importance for ecosystem management and sustainability [1].

Wildlife population assessment using traditional methods is a costly and labor-intensive process that generally relies on long-term field observations conducted by human observers. Furthermore, these approaches involve various challenges associated with climatic conditions and the physical characteristics of the terrain. Human–wildlife interactions may also expose animals to stress and disturbance, potentially posing risks to both humans and wildlife. Therefore, technological and innovative methods need to be developed to overcome these challenges associated with wildlife detection and monitoring [2].

Flamingos are colonial bird species and are considered important indicators of wetland ecosystems. Regular monitoring of flamingo populations is essential for assessing habitat quality, evaluating breeding success, and planning conservation activities. However, flamingos often occur in dense colonies comprising hundreds or even thousands of individuals, making accurate population estimation challenging. In particular, overlapping individuals, dense aggregations, and repeated counting of the same individual at different times may lead to counting errors in approaches relying solely on object detection [3–4].

To address these challenges, individual re-identification (Re-Identification, Re-ID) has emerged as an important research area. Re-ID systems aim to recognize the same individual across different images or video frames without requiring physical marking or tagging. Most existing studies utilize distinctive natural characteristics specific to individuals, such as stripes, spots, or fur patterns, to distinguish one individual from another [5–6].

However, unlike many species commonly investigated in the literature, flamingos do not possess highly distinctive individual visual patterns. Their similar body morphology, plumage coloration, and limited discriminative visual characteristics make individual re-identification in flamingos a particularly challenging research problem. Therefore, rather than relying on a single visual characteristic, multiple features, including color, texture, shape, and edge information, need to be considered jointly. Moreover, unlike terrestrial animals, the computer vision-based detection of animals in aquatic environments presents additional challenges. Reflections on the water surface, mirror-like images of animals, and highly dynamic background variations can complicate both animal detection and foreground–background separation, thereby making subsequent Re-ID more challenging. In addition, most existing studies have focused on species with distinctive individual characteristics, such as stripes, spots, or fur patterns [7]. Determining which visual features provide greater discriminative capability for species with highly similar appearances and incorporating such information into the re-identification process therefore remains an important research problem.

In this study, we aim to monitor flamingo populations and estimate unique individuals from unmanned aerial vehicle (UAV) imagery. In the first stage, flamingos are detected using a YOLO-based object detection model and tracked using tracking algorithms.

In the second stage, the detected flamingos are separated from the background using the High-Quality Segment Anything Model (HQ-SAM). In the final stage, Re-ID is performed using deep

learning-based methods to re-identify individuals whose morphological representations have been isolated from the background through segmentation. The proposed framework thus aims to estimate unique individuals in wildlife populations without requiring physical marking or tagging.

The main contributions of this study can be summarized as follows:

- A flamingo dataset comprising imagery acquired at different flight altitudes, locations, and times, including challenging environmental and visual conditions, was used in the experiments and is intended to be made available to the research community.
- To the best of our knowledge, this study represents the first investigation to apply Re-Identification (Re-ID)-based individual identification specifically to flamingos.
- The effects of UAV flight altitude on both object detection and Re-ID performance were systematically investigated.
- The proposed flamingo-focused framework is expected to provide a methodological basis for future population-counting studies involving other waterbird species.

The experimental results demonstrate that the proposed approach achieves high detection performance in flamingo population monitoring and provides a feasible method for estimating unique individuals. These findings indicate the potential of integrating UAV-based detection, segmentation, tracking, and Re-ID techniques to support non-invasive wildlife population monitoring and reduce the risk of duplicate counting.

2. RELATED WORK

Researchers have conducted studies on wildlife population detection using satellite imagery to monitor wildlife over large areas [8–9]. However, satellite imagery is not widely preferred in wildlife monitoring projects due to disadvantages such as the difficulty of detecting small animals caused by resolution limitations, vulnerability to weather conditions, restricted access to up-to-date imagery, and high costs [10]. This has increased the popularity of unmanned aerial vehicle (drone) and artificial intelligence-based wildlife monitoring approaches [11]. Recent drone-based studies have mainly focused on the systematic analysis of animal behavior [12], real-time animal detection [13], and animal health detection and monitoring [14]. Deep learning algorithms are considered the most effective method for wildlife recognition, counting, and classification from images [15]. These algorithms are based on the principle of feature extraction from images using Convolutional Neural Networks (CNNs) [16]. Fast R-CNN [17] and YOLO models [18] are widely used for real-time object detection in wildlife imagery. At present, YOLO models have become indispensable deep learning algorithms for object detection due to their high detection speed (fps), suitability for drone-based applications, and high accuracy [19].

Most studies in the literature focus solely on the object detection problem. Although object detection can determine the locations of individuals within an image, it cannot prevent the repeated counting of the same individual across different frames. In the context of computer vision, Re-Identification (Re-ID) refers to the recognition of the same individual detected at different times and locations. Particularly for species living in dense colonies, estimating population size without applying Re-ID may lead to significant errors. Therefore, in recent years, the problem of individual re-identification (Re-ID) has become an important research area in wildlife studies [20].

The problem of animal re-identification is a challenging task that is affected by factors such as seasonal variations, illumination conditions, the dynamic geometric structure of animals, background complexity, image quality, growth, molting, aging, and disease-related changes in color and shape, often resulting in unsuccessful outcomes [21–25]. Due to these challenges, Re-ID studies in wildlife detection have been conducted on a limited number of species, including sea turtles [26], polar bears [27], salamanders [28], cattle [29], tigers [30], and wasps [31]. These studies provide a universal Global ID following wildlife detection and enable the recognition of the same individual in images acquired at different times [32]. CNN-based feature extraction models [33], local feature extraction methods based on patterns, color, and shape [34], and keypoint-based feature extraction approaches [35] are commonly used for Re-ID applications.

3. MATERIALS AND METHODS

3.1. Flamingo Detection and Re-ID Algorithm

Figure 1 presents the flowchart of the proposed framework, which was developed specifically for flamingos and has the potential to serve as a basis for future population-counting studies of other waterbird species. The ByteTrack multi-object tracking algorithm was employed to track detected flamingo individuals across consecutive frames and assign a unique tracking identity (track ID) to each individual. YOLO-based object detection and tracking were applied to UAV-acquired video sequences to detect individuals within the same scene while minimizing repeated counting. To uniquely count individuals within a given scene, each detected object must be associated across consecutive frames and assigned a persistent track ID [36].

For the subsequent feature extraction stage, however, the image region corresponding to an object assigned a track ID was required to have a minimum size of 30 pixels to preserve sufficient morphological information. Although YOLO enables real-time object detection, objects become progressively smaller and contain less discriminative visual information as their distance from the imaging platform increases. To address this limitation, specialized YOLO-based approaches optimized for small-object detection have been proposed in the literature [37–39]. Based on observations obtained from the field data, track ID validation in this study was performed within a Region of Interest (ROI) corresponding to the experimental configuration that achieved the highest performance, namely a flight altitude of 30 m and an approximately 1-ha observation area.

In addition, because flamingos may appear in highly dense and overlapping groups in some scenes, a duplicate-control procedure was incorporated to reduce overlapping detections and prevent multiple track IDs from being assigned to the same individual. During duplicate control, the degree of overlap between detected bounding boxes was quantified using the Intersection over Union (IoU) metric [40], while the distances between the centers of the corresponding bounding boxes were also calculated. Furthermore, an individual assigned a track ID was required to be consistently detected for at least five consecutive frames before being considered a validated track. Once all validation criteria were successfully satisfied, the assigned track ID was confirmed and the corresponding individual was transferred to the second stage of the proposed framework.

Since the entire procedure was evaluated using video sequences of at least one minute in duration, computational performance was assessed by measuring the frame processing rate in frames per second (fps) for both the detection model and the complete proposed algorithm.

During the flamingo detection and tracking stage, the track ID assigned to each detected individual in every frame was recorded together with its confidence score, bounding-box dimensions, and coordinate information. These data were subsequently stored and reported for further analysis.

In the second stage of the proposed algorithm, flamingos that had been successfully detected and assigned validated track IDs were separated from the background using the HQ-SAM method. Previous studies have demonstrated that the original SAM model may produce insufficient mask quality, particularly for objects with complex structures and fine boundary details. In contrast, HQ-SAM can generate more detailed and higher-quality segmentation masks by incorporating a *High-Quality Output Token* and fusing early and final ViT features [41].

For Re-ID-based individual identification, four deep learning architectures originally developed and widely used for person re-identification were transferred to the task of flamingo individual identification and comparatively evaluated. These methods included BoT-R50 (Bag of Tricks) [42], AGW-R50 (Attention Generalized Mean Pooling with Weighted Regularization) [43], MGN (Multiple Granularity Network) [44], and OSNet (Omni-Scale Network) [45]. Through this transfer-learning-based framework, the applicability and comparative performance of established person Re-ID architectures for distinguishing visually similar flamingo individuals were systematically investigated.

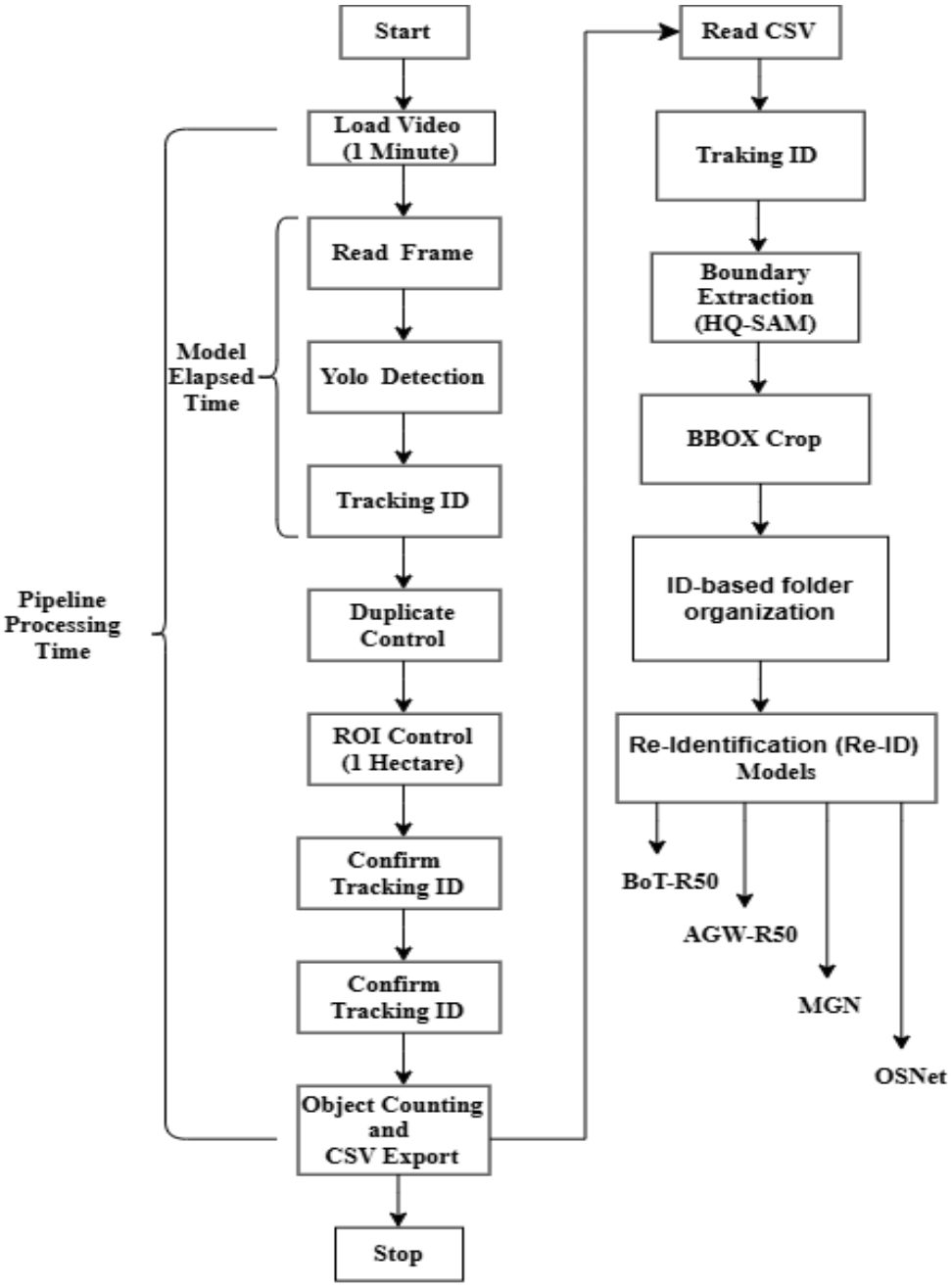

Fig. 1: Flowchart of the detection and Re-ID-based individual identification framework.

3.2. Object Detection and Model Training Using YOLO

In this study, the YOLOv8m model [45–46], which has been widely adopted due to its balance between computational efficiency and detection performance, and all variants of YOLO26 were comparatively evaluated for flamingo detection. While YOLOv8m was selected as a stable and widely used baseline model with high detection performance [47], YOLO26 was included in the comparison because of its more recent architectural improvements and enhancements for small object detection. Thus, the effects of different YOLO architectures on detection performance, processing speed, and small-object sensitivity were analyzed for unique flamingo detection.

Figure 2 presents the object detection workflow of both YOLOv8 and YOLO26. The general architecture of YOLO consists of three main components: Backbone, Neck, and Head. In the Backbone, convolutional filters slide over the input image to generate feature maps. One of the key innovations of YOLOv8 compared with previous models is the C2f (Cross Stage Partial Feature Fusion) block. As the image passes through deeper layers, particularly small objects may disappear, resulting in information loss. To address this issue, the feature map produced after convolution is divided into two parts. One half is preserved without modification to retain fine details, while the other half undergoes additional convolution operations to extract deeper features. The SPPF (Spatial Pyramid Pooling Fast) module performs multi-scale feature extraction by applying max-pooling operations at different scales. In YOLO26, the C3k2 module is used instead of the C2f module. In addition, a C2PSA (Cross Stage Partial Position-Sensitive Attention) module is applied after the SPPF stage. The C3k2 module is an enhanced feature extraction block developed for YOLO26. It improves deep feature extraction through multiple bottleneck transformations and feature fusion operations. As a result, more discriminative deep features are extracted, enabling better learning of small-object details. The C2PSA (Cross Stage Partial Position-Sensitive Attention) module employs an attention mechanism to emphasize important spatial regions while reducing background noise. This structure improves object detection performance by providing a more effective representation of small objects and complex scenes.

The Neck component of YOLO is responsible for combining feature maps from different levels of the backbone and forwarding them to the detection head. In the Upsample block, the feature map is enlarged to recover fine details of small objects. In the Concat block, feature maps obtained from different layers are merged. The primary purpose of using three different detection layers in the YOLO head is to perform multi-scale object detection. High-resolution feature maps are used for detecting small objects, whereas low-resolution feature maps provide semantic representations of larger objects. This architecture enables objects of different sizes to be detected more effectively within the same network. The recently developed YOLO26 model has gained attention in the literature due to its success in small object detection and its suitability for deployment on low-performance devices [48].

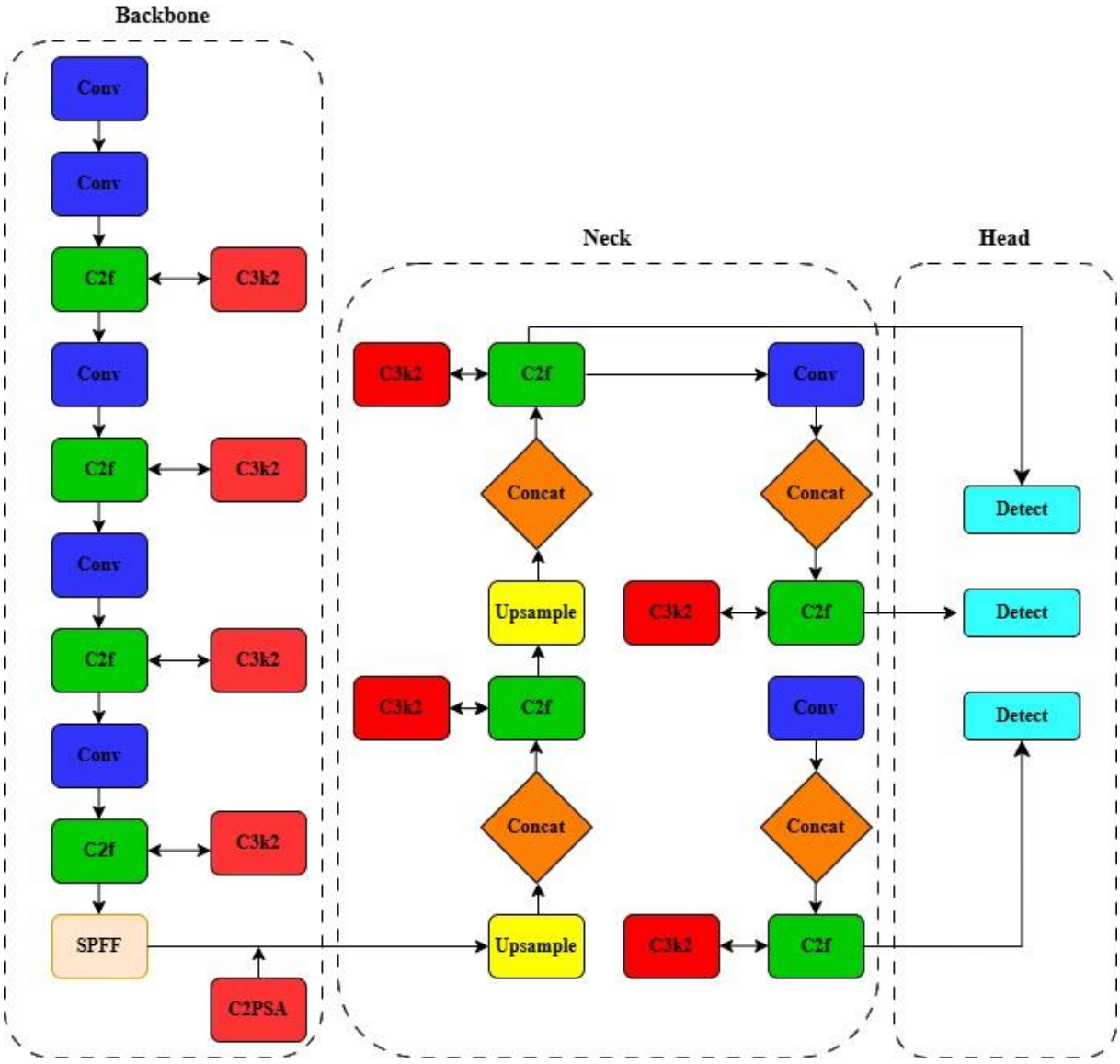

Fig. 2: YOLOv8–YOLO26 Object Detection Flow Diagram.

For flamingo detection model training, if the model correctly detects a flamingo, it is classified as TP (True Positive); if the model detects another object in the background instead of a flamingo, it is classified as FP (False Positive); if the model fails to detect a flamingo, it is classified as FN (False Negative); and if the model correctly identifies the absence of a flamingo in the background, it is classified as TN (True Negative). Therefore, the performance of the model is evaluated using the metrics Recall (1), Precision (2), Accuracy (3), AP (Average Precision) (4), and mAP (Mean Average Precision) (5), as defined by the equations given below [49].

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3}$$

In Equation (4), AP indicates how well the model performs at different recall levels and represents the area under the precision–recall curve. AP expresses the detection performance of a single class. In Equation (5), the mAP metric represents the mean of the AP values across all classes, where *n* denotes the number of classes.

$$AP = \int_0^1 Precision(Recall) d(Recall) \tag{4}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{5}$$

To prevent overlapping detections of distant flamingos in the image scene and to avoid assigning multiple detection boxes to the same flamingo, duplicate suppression and the Intersection over Union (IoU) metric were incorporated into the algorithm. The overlap ratio between two bounding boxes is calculated using the equation given in Equation (6).

$$iou = \frac{area\ of\ overlap}{area\ of\ union} \tag{6}$$

The equation presented in Equation (7) was used to calculate the Counting Accuracy as a performance metric for evaluating the counting performance of the models. The ground-truth number of flamingos is denoted by $N_{\text{ground truth}}$, whereas N_{confirm} represents the number of flamingos obtained by the proposed algorithm after duplicate removal and filtering procedures.

$$\text{Counting Accuracy}(\%) = \left(1 - \frac{|(N_{\text{confirm}}) - N_{\text{ground truth}}|}{N_{\text{ground truth}}}\right) x100 \tag{7}$$

The hyperparameters used for training both the YOLOv8m and YOLO26 models on all visual datasets employed for flamingo detection are presented in Table 1.

Table 1: Experimental Hyperparameter Settings

Hyperparameter	**Value**
Initial learning rate	0.01
Input image size	1280x1280
Batch size	2
Number of epochs	60
Optimizer	AdamW
Weight decay	0.0005
Momentum	0.937

3.3. Segmentation and Re-ID Process

The High-Quality Segment Anything Model (HQ-SAM) was employed to accurately separate the detected flamingos from the background and obtain segmentation masks representing their morphological characteristics. Figure 3 illustrates a close-up scene of flamingos after the masking process. Accurate foreground–background separation is particularly important, as the quality of segmentation directly affects the performance of deep learning-based Re-ID methods for individual identification.

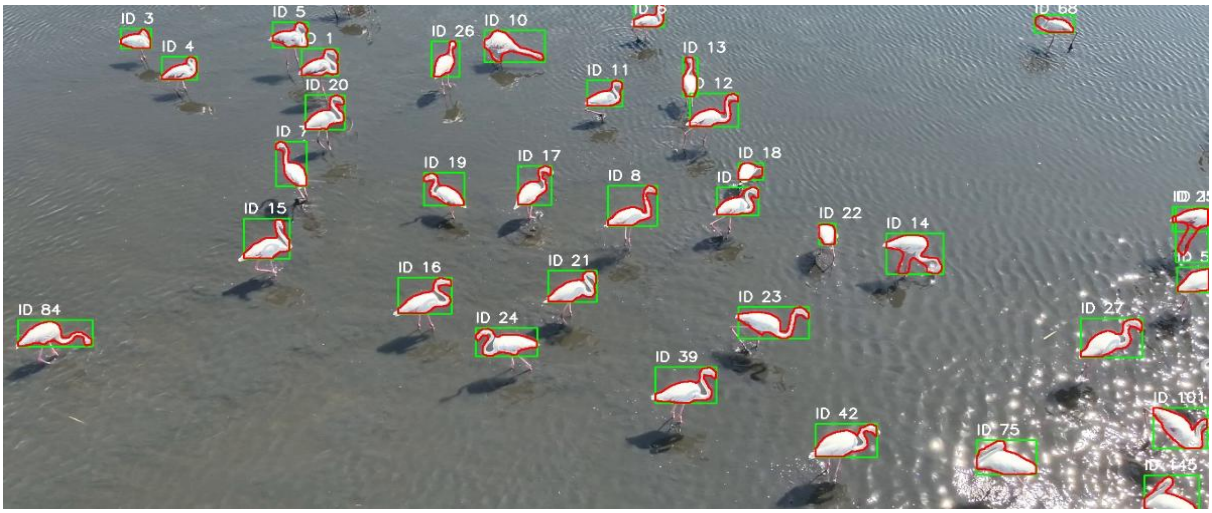

Fig. 3: Masking with HQ-SAM after detection using the YOLOv8m model.

Following the masking process, the deep learning-based BoT-R50, AGW-R50, MGN, and OSNet methods were comparatively employed for individual flamingo identification. The parameters presented in Table 2 were used during model training for feature extraction.

Table 2. Re-ID training hyperparameter settings.

Hyperparameter	Value
Initial learning rate	0.00035
Weight decay	0.0005
Batch size	32
Number of epochs	100
Train_ratio	0.60
Query_ratio	0.20
Gallery_ratio	0.20
Image_height	256
Image_width	128

Rank-1, Rank-5, and mean Average Precision (mAP) metrics were used to evaluate Re-ID performance. Rank-1 represents the model's ability to distinguish a query flamingo from other individuals and retrieve the correct identity as the top-ranked match. Rank-5 indicates the proportion of queries for which the correct flamingo identity is retrieved among the top five candidates ranked by the model. Mean Average Precision (mAP) is calculated as the mean of the Average Precision (AP) values across all query flamingos, where AP considers the ranking positions of correct identity matches within the gallery. This metric evaluates the overall ability of the Re-ID model to distinguish images of the same individual from those of different individuals and rank the correct matches at higher positions in the retrieval results.

In Equation (8), (G) denotes the total number of samples in the gallery, N_q represents the number of gallery samples sharing the same identity as the query, $P_q(k)$ denotes the precision calculated up to the (k)-th ranked result, and $rel_q(k)$ indicates whether the sample at rank (k) is a correct match to the query identity.

$$AP(q) = \frac{1}{N_q} \sum_{k=1}^{G} P_q(k) \times rel_q(k) \tag{8}$$

In Equation (9), the mean performance is calculated using the AP values obtained from Equation (8). Here, (Q) denotes the total number of query samples used to calculate the mean Average Precision (mAP).

$$mAP = \frac{1}{Q} \sum_{q=1}^{Q} AP_q \tag{9}$$

4. RESULT AND DISCUSSION

The image data used in the flamingo population detection study were acquired in the İzmir Bird Paradise under different time periods, locations, and weather conditions using a DJI Mavic 3 Enterprise series drone. As a result, the collected imagery was enriched in terms of both flamingo and background diversity. Unlike terrestrial imaging, this environment also introduces physical challenges such as waves, reflections, and refractions occurring on the water surface. The training, testing, and performance evaluation of the unique flamingo population counting algorithm were conducted in the Google Colab environment using an L4 GPU.

4.1. Impact of Altitude on Detection Performance

To investigate the effect of altitude on YOLO-based flamingo detection performance, both YOLOv8m and YOLO26m models were trained and evaluated comparatively. The models were trained using 5,500 images acquired from flights conducted below 15 m and subsequently tested on 550 images captured at altitudes of 15, 20, 25, and 30 m from the same location, which had not been previously seen by the models. The dataset was divided into 90% training and 10% validation data, resulting in a total of 5,500 images for model development. Testing was performed on a cloudy day to increase the difficulty of flamingo detection. Cloudy conditions reduce the contrast between the water surface and flamingos by causing the water color to appear duller and more similar to the color of flamingos, thereby making detection more challenging. The training results obtained after 60 epochs using the hyperparameters listed in Table 1 are presented in Table 3. No substantial difference was observed between YOLOv8m and YOLO26m in terms of training and validation performance. Considering precision, recall, and mAP50 metrics, YOLOv8m achieved slightly better results in both training and validation stages.

Table 3: Performance Results of the Trained Models Used to Investigate the Effect of Altitude on Detection Performance.

Model		Images	Instances	Precision	Recall	mAP50	mAP50-95
YOLOv8m	Train	4950	913502	0.984	0.864	0.922	0.785
	validation	550	100301	0.982	0.872	0.925	0.762
YOLO26m	Train	4950	913502	0.983	0.863	0.92	0.785
	validation	550	100301	0.98	0.87	0.923	0.781

Figure 4 illustrates the variation of the performance metrics of the YOLOv8m model, which was trained for the altitude experiments and achieved slightly better results in terms of precision, recall, and mAP50, as a function of the number of epochs. The simultaneous decrease of both training and validation losses indicates that the training process was completed without exhibiting overfitting.

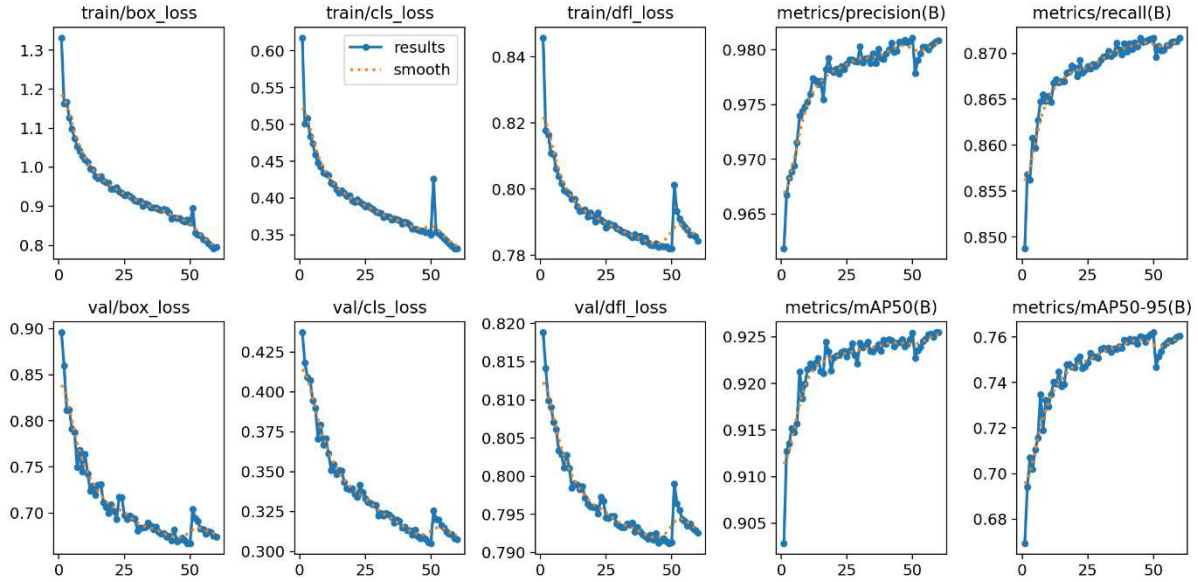

Fig. 4: YOLOv8m Training Performance Curves for Altitude-Based Performance Evaluation.

Table 4 presents the altitude-dependent test performance results of the YOLOv8m and YOLO26 models trained for the altitude experiments. In drone imagery, it has been reported in the literature that increasing flight altitude reduces object size, which generally leads to lower precision, recall, and mAP values and makes small-object detection more challenging for YOLO-based models [55–56]. While this observation remains valid in scenes with little or no object overlap, the opposite trend can be observed in scenes characterized by high individual density and significant occlusion. As shown in Table 3, the test results generally exhibit improvements in precision, recall, and mAP50 with increasing altitude, with the best performance achieved at an altitude of 30 m. In the test evaluations, YOLO26m demonstrated slightly superior performance at 30 m compared with YOLOv8m.

Table 4: Results of the Effect of Altitude on Flamingo Detection Performance.

Model	Altitude	Images	Instances	Precision	Recall	mAP50	mAP50-95
YOLOv8m	15m	550	120505	0.989	0.937	0.934	0.732
	20m	550	226295	0.987	0.944	0.944	0.687
	25m	550	183212	0.991	0.94	0.944	0.712
	30m	550	177294	0.992	0.955	0.954	0.668
YOLO26m	15m	550	120505	0.991	0.923	0.924	0.727
	20m	550	226295	0.986	0.945	0.944	0.691
	25m	550	183212	0.992	0.927	0.924	0.716
	30m	550	177294	0.992	0.956	0.955	0.714

Figure 5 presents a prediction result obtained using the YOLO26m model, which achieved an average test performance of 92.4% at an altitude of 15 m, on a sample scene containing overlapping flamingos. As shown in the image, a significant proportion of the flamingos are densely clustered at this altitude, causing the boundaries between individuals to become indistinct. As a result, the trained model assigns more detection boxes to the flamingo cluster than the actual number of individuals, leading to inaccurate population estimates. Therefore, in dense scenes with substantial overlap, increasing the flight altitude can improve detection performance by reducing occlusion between individuals, even though object resolution decreases. The obtained results indicate that detection performance depends not only on object pixel resolution but also on object density and the degree of occlusion within the scene.

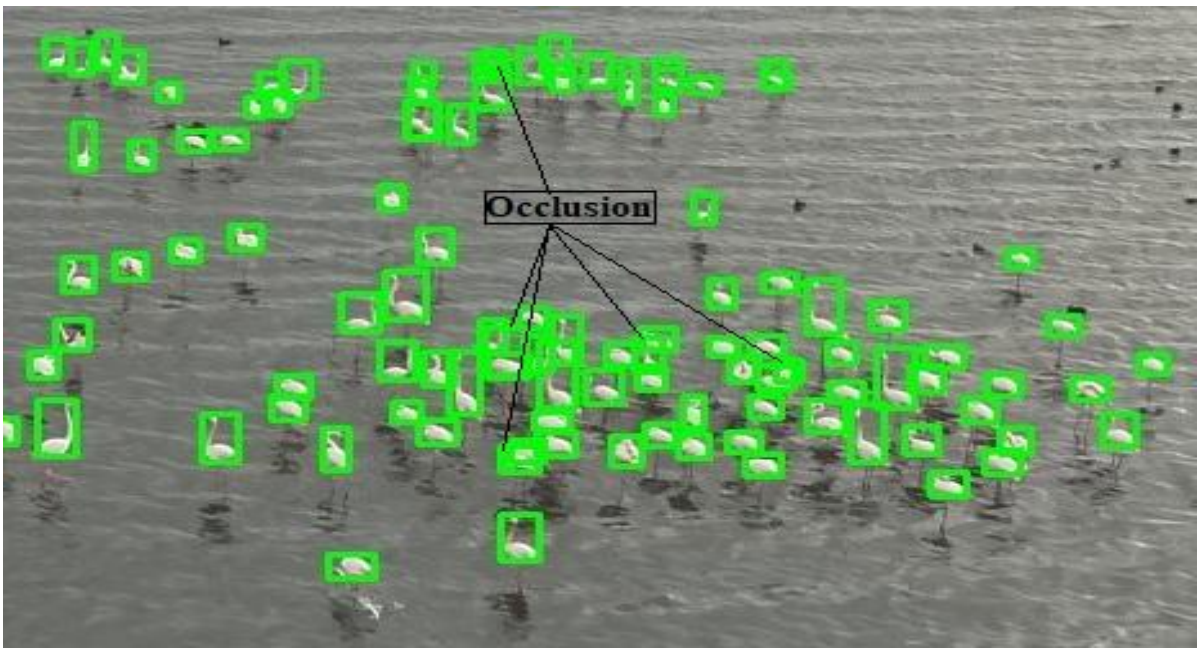

Fig. 5: YOLO26m Prediction on an Overlapping Flamingo Scene at 15 m Altitude.

4.2. Overall Detection Performance

In this study, the performance metrics and generalization capabilities of YOLOv8m and all YOLO26 variants were evaluated using a total of 6,867 images acquired at different altitudes ranging from 5 m to 30 m, different times, and different locations. The dataset was divided into 80% training, 10% validation, and 10% testing subsets. In addition, the effects of water-surface reflections, waves, and glare on population counting performance were investigated.

Table 5 presents the performance results obtained for flamingo detection. The YOLOv8m model achieved better recall and mAP50 performance. In addition, to compare frame processing speed (fps), the algorithm was executed on an L4 GPU in the Google Colab environment. To ensure a fair comparison among models, fps measurements were performed on the same scene acquired at an altitude of 30 m. The YOLOv8m model achieved superior performance with a processing speed of 14.51 fps. Despite the NMS-free architecture of YOLO26, YOLOv8m may operate faster on certain scenes when executed on a powerful GPU [57]. This is because Non-Maximum Suppression (NMS) can be processed very efficiently on high-performance hardware, reducing the fps advantage that may be observed on less powerful systems.

Table 5: Comparative Evaluation of Overall Flamingo Population Detection Performance.

Model		Images	Instances	Precision	Recall	mAP50	mAP50-95	Fps
YOLOv8m	Train	5493	1127064	0.985	0.873	0.927	0.774	14.51
	Validation	686	134727	0.98	0.885	0.935	0.738	
	Test	688	148280	0.981	0.961	0.964	0.763	
YOLO26s	Train	5493	1127064	0.98	0.869	0.924	0.764	13.62
	Validation	686	134727	0.979	0.881	0.932	0.757	
	Test	688	148280	0.982	0.938	0.933	0.736	
YOLO26m	Train	5493	1127064	0.983	0.873	0.927	0.774	13.54
	Validation	686	134727	0.982	0.885	0.934	0.767	
	Test	688	148280	0.983	0.95	0.953	0.755	
YOLO26l	Train	5493	1127064	0.984	0.873	0.927	0.774	12.24
	Validation	686	134727	0.982	0.886	0.934	0.767	
	Test	688	148280	0.983	0.952	0.953	0.757	
YOLO26x	Train	5493	1127064	0.981	0.871	0.924	0.758	12.08
	Validation	686	134727	0.979	0.883	0.932	0.752	
	Test	688	148280	0.98	0.948	0.943	0.739	

Figure 6 presents the performance curves of the YOLOv8m model, which achieved better recall and mAP results in the test evaluations in terms of overall detection performance. The simultaneous decrease in both training and validation losses indicates that the model did not exhibit overfitting during training.

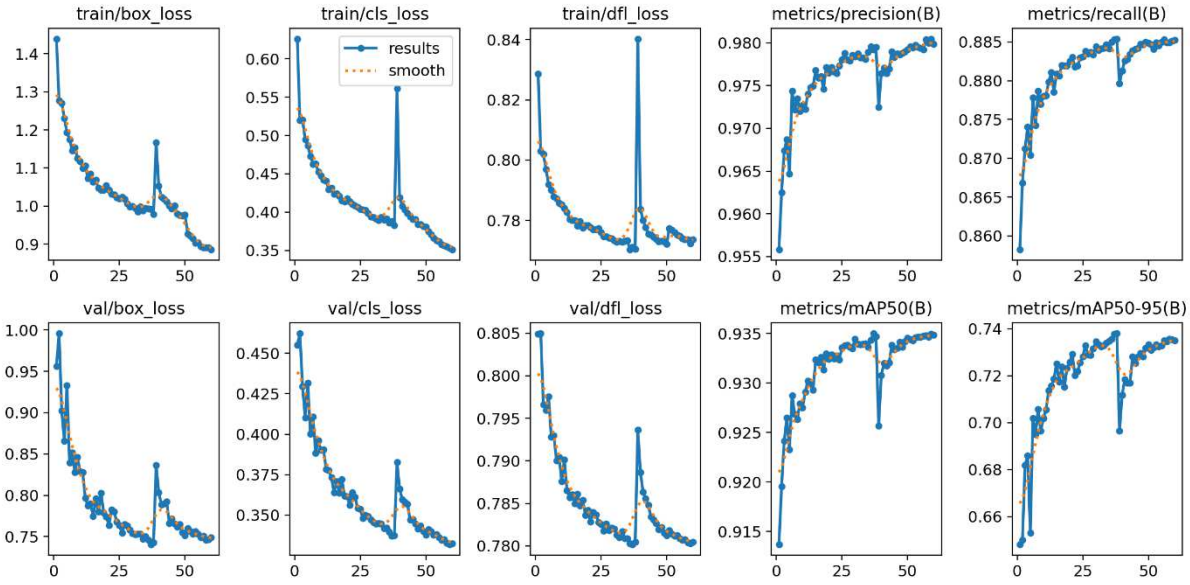

Fig. 6: YOLOv8m Performance Curves for Overall Flamingo Detection.

Figure 7 shows the flamingo population counting interface. To enable feature extraction for unique individual identification, an area of approximately 1 hectare was defined as the Region of Interest (ROI), where shape, texture, color, and edge features of flamingos could be reliably extracted.

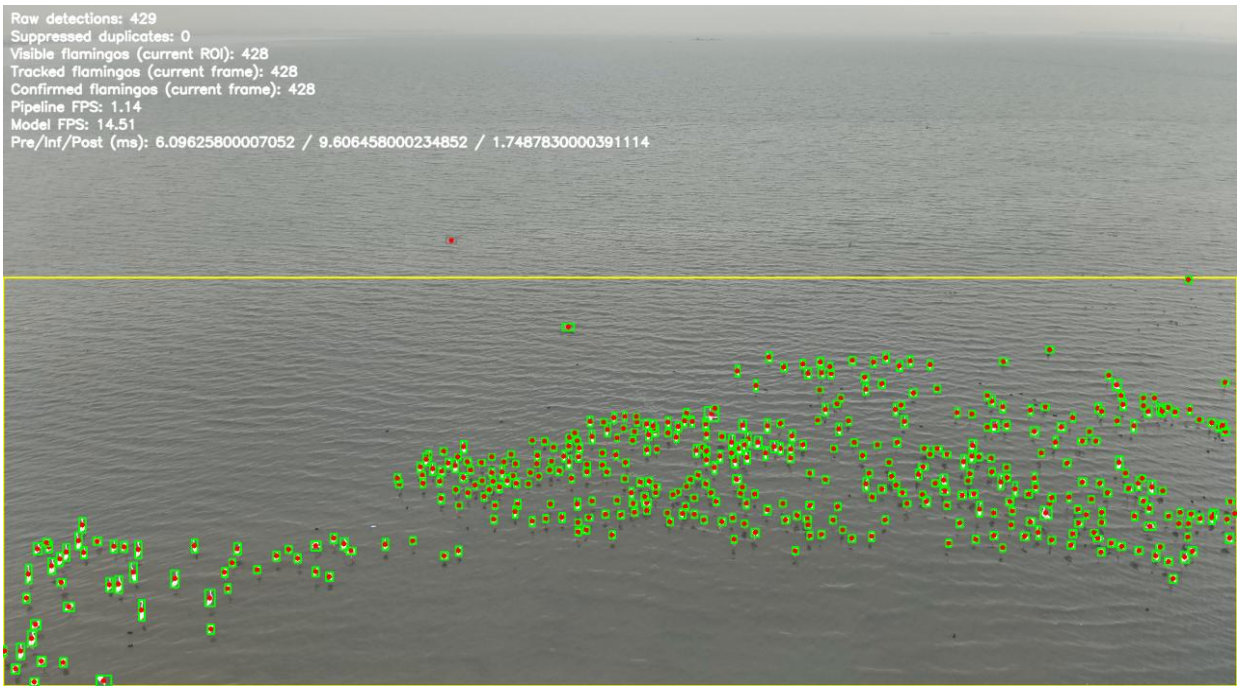

Fig. 7: Overall Flamingo Detection Results Obtained with the YOLOv8m Model at an Altitude of 30 m.

As shown in Figure 8, the proposed algorithm successfully performed population detection even in what is likely the most challenging scene, captured at an altitude of 5.3 m, where the poorest viewing angle was obtained and the highest levels of occlusion, water reflections, and shadowing were present. Due to the substantial overlap among individuals, duplicate suppression effectively eliminated excessive detection boxes, while shadows and image reflections were successfully excluded from the detection results.



Fig. 8: Overall Flamingo Detection Results Obtained with the YOLO26l Model at an Altitude of 5.3 m.

Table 6 presents the flamingo counting performance obtained using the YOLOv8m model under combinations of different flight altitudes, acquisition days, time periods, and weather conditions. Algorithmic measures were implemented to prevent duplicate counting. In this context, Suppressed Duplicates represents the bounding boxes identified by the algorithm as duplicate detections and subsequently excluded from the final count. The Confirmed Flamingos value represents the final number of validated flamingo individuals after duplicate detections have been removed.

According to the counting results, the highest counting accuracy was achieved at a flight altitude of 30 m. As the flight altitude decreased, the frequency of overlap among flamingo individuals increased, resulting in a higher number of duplicate detections. These findings indicate that flight altitude influences the degree of individual overlap and, consequently, the duplicate suppression and counting performance of the proposed algorithm.

Table 6. Flamingo counting performance results using the YOLOv8m model.

Date	Altitude (metre)	$N_{ground\ truth}$	$N_{confirm}$	Missed Count	Duplicate Count	Counting Accuracy(%)
Day-1 Cloudy	15	247	240	6	1	97.17
	20	540	535	4	1	99.08
	25	330	324	6	0	98.14
	30	429	426	3	0	99.29
Day-2 Sunny	5.3	365	361	2	2	98.90
	10	168	159	7	2	94.64

4.3. Effect of Flight Altitude on Re-ID Performance

Figure 9 shows the masked images obtained using HQ-SAM segmentation for the flamingos detected in Figure 8. Accurately extracting the complete silhouette of each flamingo, fully separating the individual from the background, and ensuring that the segmentation mask does not extend beyond the corresponding detection bounding box are important for preserving the discriminative characteristics of the features used in the subsequent Re-ID process.



Fig. 9: HQ-SAM-Based Segmentation Results of the Flamingos Detected in Figure 8.

Figure 10 shows a close-range scene acquired for feature extraction, in which HQ-SAM segmentation was applied to the detected flamingos.



Fig. 10: Close-Up Scene After HQ-SAM Segmentation.

Table 7 presents the HQ-SAM segmentation and Re-ID dataset statistics at different UAV flight altitudes. To investigate the effect of flight altitude on Re-ID performance, flamingo video sequences acquired at altitudes of 15, 20, 25, and 30 m were used. Each video sequence had a duration of five minutes. Accordingly, the Track ID count represents the total number of tracking identities assigned throughout all frames of each video sequence.

The Invalid ID count represents the number of flamingo identities that could not be included in the Re-ID process because they failed to satisfy the dataset construction criteria. Such exclusions resulted from factors including corrupted or inadequate image crops, unsuccessful HQ-SAM segmentation, and the absence of valid images suitable for Re-ID. In contrast, the Dataset ID count represents the number of individual identities that passed the quality-control procedure and were ultimately included in Re-ID training and evaluation.

To ensure sufficient visual information for individual re-identification, each Track ID was required to contain at least five valid HQ-SAM-segmented images. Re-ID performance cannot be reliably evaluated using only a single image of an individual because no additional observation of the same identity would be available for matching. In particular, at least two images of the same individual are required for that identity to be represented in both the query and gallery sets during Re-ID evaluation. Furthermore, multiple observations per identity are required for an appropriate train/query/gallery partition. Identities represented by very few images may also introduce class imbalance and excessive repeated sampling during model training.

As shown in Table 7, HQ-SAM segmentation was not applied to every detection across all video frames. Instead, a maximum of 20 images was selected from each flamingo Track ID, with consecutive selected images separated by at least 15 frames. Consequently, the number of frames processed by HQ-SAM varied according to these selection criteria rather than the total

number of flamingo detections. Following segmentation and subsequent quality-control procedures, a substantial reduction in the number of Eligible Re-ID samples was observed.

Table 7: HQ-SAM Segmentation and Re-ID Dataset Statistics.

Altitude (m)	Total Frames	Total detection	Number of HQ-SAM Frames	Track ID	False ID	Data ID	Eligible Re-ID
15	9019	529327	11510	831	55	776	661
20	9476	706526	14239	824	128	696	511
25	9019	413466	10037	777	148	629	415
30	9017	493163	9779	549	101	448	277

As shown in Table 8, the experimental Re-ID analyses demonstrated that OSNet achieved the highest performance among the evaluated methods. At a flight altitude of 30 m, OSNet achieved a Rank-1 accuracy of 91.75%, a Rank-5 accuracy of 93.97%, and an mAP of 72.49% for individual matching in flamingo scenes.

The improvement in Re-ID performance observed with increasing flight altitude may be associated with the more homogeneous top-down views of flamingos obtained at higher altitudes and the reduced pose and perspective variations across images. Previous studies in the Re-ID literature have demonstrated that variations in viewpoint and pose can substantially affect the visual representation of the same individual and consequently reduce identity-matching performance. In particular, different camera viewpoints and body poses may result in considerable appearance variations for the same individual, thereby making the Re-ID task more challenging [54–55].

Table 8. Comparison of Re-ID model performance.

Altitude (m)	Model	Rank-1 (%)	Rank-5 (%)	mAP (%)
15	**BoT-R50**	52.54	69.56	47.15
	AGW-R50	21.90	37.36	18.83
	MGN	58.24	73.42	52.74
	OSNet	74.15	87.72	51.69
20	**BoT-R50**	65.53	82.45	57.12
	AGW-R50	24.07	44.40	19.70
	MGN	70.14	84.27	60.74
	OSNet	88.48	95.16	59.27
25	**BoT-R50**	65.39	78.39	59.37
	AGW-R50	27.92	46.37	25.38
	MGN	71.61	82.89	64.39
	OSNet	81.22	86.71	60.72
30	**BoT-R50**	76.70	83.03	70.60
	AGW-R50	42.23	64.08	38.59
	MGN	78.96	89.81	71.94
	OSnet_x1_0	**91.75**	**93.97**	**72.49**

5. CONCLUSION

In this study, a Re-ID-based individual identification framework was developed for flamingos, with the potential to provide a methodological basis for the unique counting of other waterbird species in future population monitoring studies. The proposed approach was evaluated using real-world field data, and the experimental findings were presented to demonstrate its applicability under natural environmental conditions.

The effects of UAV flight altitude on both object detection and Re-ID-based individual identification performance were investigated for flamingos. For the object detection stage, YOLOv8m and YOLO26 models were compared in terms of detection performance and processing speed. For the Re-ID stage, the individual matching performances of the deep learning-based BoT-R50, AGW-R50, MGN, and OSNet methods were comparatively evaluated. The highest flamingo detection performance was achieved using YOLOv8m on imagery acquired at a flight altitude of 30 m over an approximately 1-ha study area. The YOLOv8m model achieved an average test performance (mAP@0.50) of 96.4% for flamingo detection. In the Re-ID-based individual identification task, OSNet achieved the highest performance, with Rank-1, Rank-5, and mAP scores of 91.75%, 93.97%, and 72.49%, respectively.

Future studies will focus on preventing duplicate counts through Re-ID-based individual identification and subsequently estimating population size based on the number of unique individuals in wildlife inventory applications. In addition, the re-identification capability of individuals under variations in acquisition time, UAV flight altitude, viewing angle, illumination, and environmental conditions will be investigated. Further research will also evaluate the cross-species generalization capability of the proposed approach by applying it to different wildlife species. Additional objectives include long-term monitoring of individual movement patterns and habitat use, as well as transforming the proposed framework into a real-time wildlife monitoring system.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the General Directorate of Nature Conservation and National Parks for providing access to the İzmir Delta Wetland–Bird Sanctuary, where the flamingo data were collected and the test flights were conducted. The authors also gratefully acknowledge the management and staff of the 4th Regional Directorate of Nature Conservation and National Parks for providing the technical facilities, capabilities, and support that contributed to the successful completion of this study.

Funding Declaration: This research received no external funding.

Data and Code Availability Statement: The datasets and algorithm implementations used in this study will be made available through a GitHub repository upon publication of the article.

6. REFERENCES

[1] Wang, S., Chen, K., Wei, Z., Yang, L., Wang, Q., Liu, M., ... & Xu, F. (2026). UAV-based deep learning for biodiversity monitoring: advances, applications, and future directions. *Ecological Informatics*, 103710.

[2] Aliane, N. (2025). Drones and AI-driven solutions for wildlife monitoring. *Drones*, *9*(7), 455.

[3] Descamps, S., Descombes, X., Bechet, A., & Zerubia, J. (2009). Flamingo detection using marked point processes for estimating the size of populations. *Traitement du Signal*, *26*(2), 95-108.

[4] Yang, S., Francis, R. J., Holding, M., & Kingsford, R. T. (2024). Aerial photography and machine learning for estimating extremely high flamingo numbers on the Makgadikgadi Pans, Botswana. *Global Ecology and Conservation*, *53*, e03011.

[5] Li, S., Li, J., Tang, H., Qian, R., & Lin, W. (2019). ATRW: A benchmark for amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*.

[6] Nepovinnykh, E., Eerola, T., & Kalviainen, H. (2020). Siamese network based pelage pattern matching for ringed seal re-identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision workshops* (pp. 25-34).

[7] Ravoor, P. C., & Sudarshan, T. (2020). Deep learning methods for multi-species animal re-identification and tracking–a survey. *Computer Science Review*, *38*, 100289.

[8] Zhao, P., Liu, S., Zhou, Y., Lynch, T., Lu, W., Zhang, T., & Yang, H. (2021). Estimating animal population size with very high-resolution satellite imagery. *Conservation Biology*, *35*(1), 316-324.

[9] Xue, Y., Wang, T., & Skidmore, A. K. (2017). Automatic counting of large mammals from very high-resolution panchromatic satellite imagery. *Remote sensing*, *9*(9), 878.

[10] Delplanque, A., Théau, J., Foucher, S., Serati, G., Durand, S., & Lejeune, P. (2024). Wildlife detection, counting and survey using satellite imagery: are we there yet?. *GIScience & Remote Sensing*, *61*(1), 2348863

[11] Aliane, N. (2025). Drones and AI-driven solutions for wildlife monitoring. *Drones*, *9*(7), 455.

[12] Pedrazzi, L., Naik, H., Sandbrook, C., Lurgi, M., Fürtbauer, I., & King, A. J. (2025). Advancing animal behaviour research using drone technology. *Animal Behaviour*, *222*, 123147.

[13] Dat, N. N., Richardson, T., Watson, M., Meier, K., Kline, J., Reid, S., ... & Burghardt, T. (2025). WildLive: Near real-time visual wildlife tracking onboard UAVs. *arXiv preprint arXiv:2504.10165*.

[14] Yaney-Keller, A., McIntosh, R. R., Clarke, R. H., & Reina, R. D. (2025). Closing the air gap: the use of drones for studying wildlife ecophysiology. *Biological Reviews*, *100*(3), 1206-1228.

[15] Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, *115*(25), E5716-E5725.

[16] Hou, J., He, Y., Yang, H., Connor, T., Gao, J., Wang, Y., ... & Zhou, S. (2020). Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*, *242*, 108414.

[17] Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., & Ye, H. (2020). Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, 364-376.

[18] Li, Y., Li, S., Du, H., Chen, L., Zhang, D., & Li, Y. (2020). YOLO-ACN: Focusing on small target and occluded object detection. *IEEE access*, *8*, 227288-227303.

[19] Raza, A., Hanif, F., & Mohammed, H. A. (2025). Analyzing the enhancement of CNN-YOLO and transformer-based architectures for real-time animal detection in complex ecological environments. *Scientific Reports*, *15*(1), 39142.

[20] Beyan, C., Tur, A. O., & Karimi, E. From Species-Specific Models to Universal Re-ID: A Survey of Animal Re-Identification. *Available at SSRN 6117620*.

[21] Li, S., Li, J., Tang, H., Qian, R., & Lin, W. (2019). ATRW: A benchmark for amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*.

[22] Nepovinnykh, E., Eerola, T., Biard, V., Mutka, P., Niemi, M., Kunnasranta, M., & Kälviäinen, H. (2022). SealID: Saimaa ringed seal re-identification dataset. *Sensors*, *22*(19), 7602.

[23] Zuerl, M., Dirauf, R., Koeferl, F., Steinlein, N., Sueskind, J., Zanca, D., ... & Eskofier, B. (2023). Polarbearvidid: A video-based re-identification benchmark dataset for polar bears. *Animals*, *13*(5), 801.

[24] Chan, J., Carrión, H., Mégret, R., Agosto-Rivera, J. L., & Giray, T. (2022, February). Honeybee Re-identification in Video: New Datasets and Impact of Self-supervision. In *VISIGRAPP (5: VISAPP)* (pp. 517-525).

[25] Jiao, B., Liu, L., Gao, L., Wu, R., Lin, G., Wang, P., & Zhang, Y. (2023). Toward re-identifying any animal. *Advances in Neural Information Processing Systems*, *36*, 40042-40053.

[26] Adam, L., Čermák, V., Papafitsoros, K., & Picek, L. (2024). SeaTurtleID2022: A long-span dataset for reliable sea turtle re-identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 7146-7156).

[27] Zuerl, M., Dirauf, R., Koeferl, F., Steinlein, N., Sueskind, J., Zanca, D., ... & Eskofier, B. (2023). Polarbearvidid: A video-based re-identification benchmark dataset for polar bears. *Animals*, *13*(5), 801.

[28] Matthé, M., Sannolo, M., Winiarski, K., Spitzen-van der Sluijs, A., Goedbloed, D., Steinfartz, S., & Stachow, U. (2017). Comparison of photo-matching algorithms commonly used for photographic capture–recapture studies. *Ecology and evolution*, *7*(15), 5861-5872.

[29] Bergamini, L., Porrello, A., Dondona, A. C., Del Negro, E., Mattioli, M., D'alterio, N., & Calderara, S. (2018, November). Multi-views embedding for cattle re-identification. In *2018 14th international conference on signal-image technology & internet-based systems (SITIS)* (pp. 184-191). IEEE.

[30] Shi, C., Liu, D., Cui, Y., Xie, J., Roberts, N. J., & Jiang, G. (2020). Amur tiger stripes: Individual identification based on deep convolutional neural network. *Integrative zoology*, *15*(6), 461-470.

[31] Borlinghaus, P., Tausch, F., & Rettenberger, L. (2023). A purely visual re-id approach for bumblebees (bombus terrestris). *Smart Agricultural Technology*, *3*, 100135.

[32] Jiao, B., Liu, L., Gao, L., Wu, R., Lin, G., Wang, P., & Zhang, Y. (2023). Toward re-identifying any animal. *Advances in Neural Information Processing Systems*, *36*, 40042-40053.

[33] Hou, S., Huang, P., Wang, Z., Liu, Y., Li, Z., Zhang, M., & Huang, Y. (2025). Openanimals: Revisiting person re-identification for animals towards better generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 14369-14379).

[34] Nepovinnykh, E., Chelak, I., Eerola, T., Immonen, V., Kälviäinen, H., Kholiavchenko, M., & Stewart, C. V. (2024). Species-agnostic patterned animal re-identification by aggregating deep local features. *International Journal of Computer Vision*, *132*(9), 4003-4018.

[35] Moskvyak, O., Maire, F., Dayoub, F., & Baktashmotlagh, M. (2021). Keypoint-aligned embeddings for image retrieval and re-identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 676-685).

[36] Adžemović, M. (2025). Deep learning-based multi-object tracking: A comprehensive survey from foundations to state-of-the-art. arXiv preprint arXiv:2506.13457.

[37] Wang, P., & Zhao, J. (2025). Sod-Yolo: enhancing Yolo-based detection of small objects in Uav imagery. arXiv preprint arXiv:2507.12727.

[38] Lu, L., He, D., Deng, Z., Liu, C., & Shen, Z. (2025, October). MASF-YOLO: an improved YOLOv11 network for small object detection on drone view. In 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2703-2708). IEEE.

[39] Han, Y., Wang, C., Luo, H., Wang, H., Chen, Z., Xia, Y., & Yun, L. (2025). LRDS-YOLO enhances small object detection in UAV aerial images with a lightweight and efficient design. Scientific Reports, 15(1), 22627.

[40] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 658-666).

[41] Ke, L., Ye, M., Danelljan, M., Tai, Y. W., Tang, C. K., & Yu, F. (2023). Segment anything in high quality. *Advances in Neural Information Processing Systems*, *36*, 29914-29934.

[42] Luo, H., Gu, Y., Liao, X., Lai, S., & Jiang, W. (2019, June). Bag of tricks and a strong baseline for deep person re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1487-1495). IEEE.

[43] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. (2021). Deep learning for person re-identification: A survey and outlook. *IEEE transactions on pattern analysis and machine intelligence*, *44*(6), 2872-2893.

[44] Wang et al. (2018), "Learning Discriminative Features with Multiple Granularities for Person Re-Identification." In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 274-282).

[45] Zhou, K., Yang, Y., Cavallaro, A., & Xiang, T. (2019, October). Omni-scale feature learning for person re-identification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3701-3711). IEEE.

[46] Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., Nepal, U., Poulose, A., ... & Karkee, M. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artificial Intelligence Review,* 58(9), 274.

[47] Pereira, G. A. (2024). Fall detection for industrial setups using YOLOv8 variants. arXiv preprint arXiv:2408.04605.

[48] Sohan, M., Sai Ram, T., & Rami Reddy, C. V. (2024). A review on yolov8 and its advancements. In *International conference on data intelligence and cognitive informatics* (pp. 529-545). Springer, Singapore.

[49] Hidayatullah, P., & Tubagus, R. YOLO26: A Comprehensive Architecture Overview and Key Improvements. arXiv 2026. *arXiv preprint arXiv:2602.14582*.

[50] Wang, Y., Ma, C., Zhao, C., Xia, H., Chen, C., & Zhang, Y. (2025). WB-YOLO: An efficient wild bat detection method for ecological monitoring in complex environments. *Engineering Applications of Artificial Intelligence*, *157*, 111232.

[51] Petso, T., Jamisola, R. S., Mpoeleng, D., & Mmereki, W. (2021, October). Individual animal and herd identification using custom YOLO v3 and v4 with images taken from a UAV camera at different altitudes. In *2021 IEEE 6th international conference on signal and image processing (ICSIP)* (pp. 33-39). IEEE.

[52] Gao, P., Zhong, D., Qi, Q., Ling, C., Qiu, C., Wang, B., ... & Gao, M. (2026). FDM-YOLO: Real-time small-target UAV wildlife detection via attention-guided cross-modality fusion. *Ecological Informatics*, 103697.

[53] Oguine, C. G., Oguine, K. J., Oguine, O. M., & Oguine, O. C. (2026). Multiscale Real-Time Object Detection in the NMS-Free Era: A Comparative Performance Evaluation of YOLOv8 and YOLO26. *arXiv preprint arXiv:2605.24831*.

[54] Cho, Y. J., & Yoon, K. J. (2016). Improving person re-identification via pose-aware multi-shot matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1354-1362).

[55] Zheng, R., Gao, C., & Sang, N. (2021). Viewpoint transform matching model for person re-identification. *Neurocomputing*, *433*, 19-27.