A digital framework for automated non-invasive waterfowl detection in Carinthia based on high resolution UAS imagery and machine learning

Gernot Paulus, Mohammad Sa'Doun, Karl-Heinrich Anders, Ulf Scherling, Werner Petutschnig, Johann Wagner, Christopher D. Lippitt

ABSTRACT

Automated waterfowl detection from uncrewed aerial system (UAS; "drones") imagery has become an important task for various environmental applications such as wildlife monitoring, nature conservation, and habitat mapping. This paper presents a digital framework for automated waterfowl detection using high-resolution UAS imagery and artificial intelligence/machine learning (ML). Several UAS missions in Brenndorf, Carinthia, Austria, were conducted simultaneously with a traditional ground-based waterfowl field survey by an experienced expert. Several data pre-processing steps were applied to optimize digital image data pipelines for the generation of high-quality ML training data. The You Only Look Once (YOLO) open-source computer vision and ML object detection model was used to detect waterfowl in the UAS imagery. A transfer learning approach from a large waterfowl study at the University of New Mexico in collaboration with the U.S. Fish and Wildlife Service was used to further improve the model's performance. Validation results showed promising performance with 80% and 83% classification accuracy on the waterfowl classes 'duck' and 'swan', respectively. Finally, a spatial projection model and a visualization approach for the ML-based detection and classification results on a map were implemented. The proposed digital framework for automated waterfowl detection provides promising results for standardization and a new paradigm for waterfowl counting to support and extend traditional wildlife monitoring.

KEYWORDS

- waterfowl surveying
- uncrewed aerial systems
- deep learning
- YOLO
- expert annotation
- digital geotransformation

Ein digitales Framework für die automatisierte, nichtinvasive Wasservogelerfassung in Kärnten basierend auf hochauflösenden UAS-Bildern und maschinellem Lernen

ZUSAMMENFASSUNG

Die automatische Erkennung von Wasservögeln anhand von Bildern unbemannter Luftfahrtsysteme (UAS) ist zu einer wichtigen Aufgabe für verschiedene Umweltanwendungen geworden, z. B. für die Überwachung von Wildtieren, den Naturschutz und die Kartierung von Lebensräumen. In diesem Beitrag wird ein digitaler Rahmen für die automatische Erkennung von Wasservögeln mit Hilfe von hochauflösenden Drohnenbildern und künstlicher Intelligenz vorgestellt. Mehrere UAS-Einsätze in Brenndorf im Bezirk Völkermarkt in Kärnten wurden zeitsynchron mit einer klassischen bodengestützten Wasservogelkartierung durch einen erfahrenen Fachmann durchgeführt. Mehrere Datenvorverarbeitungsschritte wurden angewandt, um digitale Bilddatenpipelines für die Generierung qualitativ hochwertiger KI-Trainingsdaten zu optimieren. YOLOv5, ein Modell aus der You Only Look Once (YOLO)-Familie von Computer-Vision-Modellen für die Objekterkennung, wurde verwendet. Um die Leistung des Modells weiter zu verbessern, wurde ein Transfer-Learning-Ansatz aus einer großen Wasservogel-Studie, die an der University of New Mexico im Auftrag des U.S. Fish and Wildlife Service durchgeführt wurde, übernommen und adaptiert. Die Validierungsergebnisse zeigten eine vielversprechende Leistung mit einer Klassifizierungsgenauigkeit von 80 % und 83 % auf Artenebene für „Ente" und „Schwan". Ein räumliches Projektionsmodell für die Darstellung der Ergebnisse der durch KI detektierten einzelnen Wasservogelindividuen in Kartenform zeigt die räumliche Verteilung der Wasservögel im Projektgebiet. Der vorgeschlagene digitale Rahmen für die automatische Erkennung von Wasservögeln liefert vielversprechende Ergebnisse für die Standardisierung und ein neues Paradigma für die Zählung von Wasservögeln, um die traditionelle Feldkartierung zu unterstützen und zu erweitern.

INTRODUCTION

Human survival is dependent upon biodiversity. Genes, species, and ecosystems sustain our food systems and protect us from disease and climate change. Many Earth ecosystems are stressed or ruined, and many animals, plant, and microbial species are at risk of extinction. Expected human population growth, resource extraction, and CO_2 emissions through the 2020s means that this destruction of life is likely to increase unless there is immediate, global, and sustained protection and regeneration of nature. Waterfowl serve as indicators of biodiversity and ecological health for a diverse set of ecosystems. Waterfow such as ducks, geese and swans are especially impacted by global changes as they rely on many different habitats throughout their annual migratory routes. Traditionally, researchers and volunteers worldwide count waterfowl in the field by surveying either from the ground or in more remote habitats even from crewed aircrafts. The problem with such types of surveys is that they are imprecise, as ground-based observations are limited by field of view, site accessibility and cause disturbance to wildlife when animals are approached by humans or loud, low-flying aircraft [1].

Wagner and Petutschnig [2] presented the results of a traditional ground-based waterfowl census performed in January 2021 by 85 volunteering observers in 27 counting districts at lakes and river segments in Carinthia, Austria. Based on a standardized taxonomy, the total number of observed species was aggregated for each counting district. This census is part of the International Waterbird Census [3], which is a global monitoring program consisting of an annual synchronized count of all waterbird species. The counts are organized during the non-breeding season when many species congregate in wetlands.

Automated waterfowl detection from uncrewed aerial systems (UAS) imagery is a promising opportunity for wildlife management, nature conservation, and habitat mapping [4]. Accurate detection and classification of waterfowl can provide valuable information on population dynamics, migratory patterns, and habitat use, which can aid in the development of effective conservation strategies [5]. In recent years, deep learning models have achieved state-of-the-art performance in various object detection and classification tasks, including waterfowl detection from UAS imagery [6]. One of the most widely used deep learning models for object detection is YOLO - You Only Look Once [7], which is a single-shot detector that can detect and classify objects in real-time. YOLO has undergone several iterations, with the latest version being YOLOv5 (Ultralytics Inc., Los Angeles, CA, USA) at the time of implementation of this study. In this paper, a pipeline for automated waterfowl detection from UAS imagery is presented. The pipeline consists of three key stages: (1) data capture and pre-processing; (2) machine learning (ML) model training and validation; and (3) map projection of detection results. In the data capture and pre-processing stage (1), a UAS mission was conducted in Brenndorf, Carinthia, Austria, resulting in a large number of overlapping images. Metashape (Agisoft LLC, St. Petersburg, Russia) software for photogrammetric processing of digital images was used to reduce the image overlap, and Labelbox (Labelbox Inc., San Francisco, CA, USA) facilitated data annotation. To prepare the data for training, the annotation format was converted from JSON to the required text file format, and image cropping was performed to improve the object-to-image size ratio. For the training and validation stage (2), transfer learning was applied using a label set and YOLO model weights from the University of New Mexico (UNM). The model's weights [8] served as the starting point for detecting waterfowl in Brenndorf. In the map projection stage (3), a projection model was constructed to visualize the detections on a map. By completing these three stages, we developed an effective pipeline for automated waterfowl detection from UAS imagery using YOLO. The main contribution of this paper is to critically reflect upon and provide valuable insights into the effectiveness of this ML-based approach for waterfowl detection and its further potential for applications in wildlife management, nature conservation, and habitat mapping.

MATERIALS AND METHODS

Surveying waterfowl using UAS equipped with high-resolution cameras represents a novel solution to the long-standing problems of how to efficiently survey waterfowl populations. The general method is to program a flight path for a UAS that goes above a popular spot where waterfowl gather, and have it take pictures of the flock as it flies over. UAS are smaller and quieter than crewed aircraft, so they can fly fairly low over flocks of waterfowl without stressing the birds. They can also conduct a survey more quickly than humans can, and over a much broader area [9].

In the course of this feasibility study, a new innovative workflow for non-invasive water bird counting based on high-resolution aerial drone imagery and artificial intelligence (AI) was designed and evaluated according to scientific criteria. Carinthian test sites for the prototypical implementation and validation of this approach were the Bleistätter Moor at the eastern end of Lake Ossiachersee and the ecological compensation measure of the Austrian Federal Railways (ÖBB) in Brenndorf at the Drava River.

The new proposed workflow consists of five phases (Figure 1). Phase 1 addresses the overall requirements and the conceptual study design. Phase 2 relates to image data acquisition, in which UAS missions are planned to capture high-resolution individual image data. Phase 3 provides analysis-ready data for ML such as pre-processing of images and annotation of selected images. Phase 4 relates to training and validation, which involves feeding the output of the previous two phases (images and annotations) into a ML algorithm (YOLO) with the goal of measuring the ML ability to accurately detect and classify objects. Phase 5 presents results where ML detections are visualized in the form of maps that show projected captured images and vector points indicating the detected locations and spatial distribution of waterfowl in the study area.

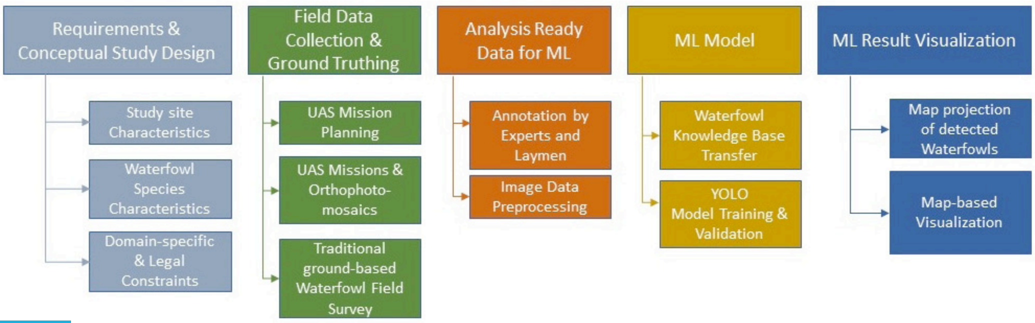

Fig. 1

Figure 1: Framework for machine learning-based waterfowl detection.
Source: own figure

Abbildung 1: *Struktur für maschinelles Lernen zur Erkennung von Wasservögeln.*
Quelle: eigene Abbildung

Requirements and Conceptual Study Design

In a first step a comprehensive requirement analysis was performed in close collaboration with waterfowl domain experts of the Carinthian Provincial Government (Dept. 8 – Environment, Nature Conservation and Climate Protection Coordination) in order to design the concept of the study. This consisted of joint meetings to select the study site, the expected waterfowl species characteristics, and any domain-specific and legal constraints that needed to be considered to safely perform the UAS missions and field survey. A key objective was that the UAS-based image capture and the traditional ground-based waterfowl survey be performed at the same date, time of the day and duration in order to be able to compare the results and exclude as much potential methodological sampling bias as possible.

The study site for this conceptual study was an ecological compensation measure and substitute biotope of the substitute biotope of the ÖBB in Brenndorf at the Drava River located in the district of Völkermarkt in Lower Carinthia. Data acquisition took place on October 6th 2022 in the eastern part of the biotope over an area of approx. 115,000 m² (Figure 2). This site was proposed by the domain experts because of the expected presence of significant numbers of waterfowl on the anticipated field survey date in October 2022. In order to perform drone missions in a responsible and safe manner, all legal constraints in terms of the common European drone regulations [10] [11] as well as domain-specific legal regulations in the context of nature conservation and protected areas (K-NBG 2019) [12] were considered.

Fig. 2

Figure 2: Bird's eye view of the Brenndorf study site (10/6/2022). Source: own figure

Abbildung 2: *Vogelperspektive auf das Untersuchungsgebiet Brenndorf (6.10.2022). Quelle: eigene Abbildung*

Field Data Collection and Ground Truthing

The collaborative data collection involved two main methodologically distinct activities, which were performed in the same defined study area simultanously, (1) the use of UAS to capture high-resolution images, and (2) the execution of a traditional, ground-based field survey for comparison purposes performed by an experienced and well-trained waterfowl domain expert.

UAS Mission Planning

Based on the conducted research and evaluations, the optimal flight altitude for UAS to capture high-resolution images for individual waterfowl identification and detection was identified empirically to be 30m above water level and present waterfowl respectively. Various flight altitudes were tested and compared in a previous internal study at the Bleistätter Moor using different UAS platforms such as the Fixed-Wing UAS BRAMOR PPK, Small Multirotor UAS DJI Phantom4 RTK, and Very Small Multirotor UAS DJI Mavic 2 Pro. The selected flight altitude of 30m was empirically determined based on the judgement of a waterfowl domain expert. This visual "*scaring off*" assessment was performed on-site directly before starting the image data capture in Brenndorf. This procedure ensured a non-invasive drone-based data capture without disturbing any waterfowl present in the study area. Such low flight altitude results in high spatial image resolution of 8 mm per pixel and therefore improved detection and recognition of individual waterfowl characteristics by domain experts as well as by ML. In order to provide an up-to-date basemap for the ground truthing field survey, a UAS mission was performed close in time two weeks before the collaborative data capture survey in order to provide a high resolution basemap orthophoto representing the actual state and context information of the environment at the study site. This additional step was necessary as available aerial images were outdated, captured at a different time of the year with different states of vegetation and water level, and were therefore unsuitable as high quality survey basemaps.

UAS Missions & Orthophotomosaics

Three missions were performed with two different UAS on two dates. Table 1 provides an overview of the UAS missions performed at the study site. All missions were performed in the open category in visual line of sight (VLOS) conditions. A DJI Phantom 4 multirotor system with a real-time-kinematic (RTK) positioning system was used for data collection, allowing cm-accuracy survey grade georeferencing of the individual captured images (Figure 3). A smaller DJI Mavic Pro 2 multirotor system was used to collect a video and several perspective views for documentation purposes.

Tab. 1

Test site	Date	Drone	Flight Altitude	Comment
Brenndorf	9/20/2022	Phantom 4 RTK	120m	Orthophoto for creating base map for expert field mapping; high image overlap
Brenndorf	10/6/2022	Phantom 4 RTK	30m	ML image capture & orthophoto; high image overlap
Brenndorf	10/6/2022	Mavic Pro 2	100m	Capture of perspective images and video for documentation purposes

Table 1: Overview of the performed UAS missions

Tabelle 1: *Übersicht über die durchgeführten UAS-Missionen*

Fig. 3

Figure 3: Start of Phantom 4 RTK multirotor UAS system at Brenndorf (10/6/2022). Source: own figure

Abbildung 3: *Start des Phantom 4 RTK Multirotor UAS Systems in Brenndorf (10/6/2022). Quelle: eigene Abbildung*

All missions were performed based on a mission plan determining the flight altitude and the overlap between collected images in terms of in-flight direction (90% front-lap) as well as between neighboring flight lines (90% side-lap). The Phantom 4 RTK performs the mission automatically based on the defined GNSS waypoints of the mission plan. Once a mission plan is defined, it can be stored and reproduced to guarantee compliant UAS mission parameters in case of multiple acquisitions in the same study site, e.g. for monitoring and change detection purposes.

Based on the individual UAS images captured by the Phantom 4 RTK, Metashape was used to photogrammetrically process individual digital images to generate high-resolution 3D point clouds, a digital surface model, and a high-resolution orthophotomosaic. The orthophotomosaic resulting from the overview mapping mission prior to the collaborative data collection was provided as an analogue basemap to domain experts and was used for the ground truthing field survey. The orthophotomosaic derived from the same day as the collaborative data collection was performed and was used as the most up-to-date base map for visualizing the ML results.

Traditional ground-based waterfowl field survey

Traditional waterfowl field mapping was performed by a domain expert in order to validate and assess the results of the ML detection and classification process. A highly experienced field expert, following a field data protocol and moving along the water's edge, recorded the location of all observed and identified waterfowl on the high-resolution, up-to-date analogue basemap. An important issue was the simultaneous field survey within the same time window as the UAS missions and the mapping of the observed waterfowl from the ground based on visual sight observations with the support of professional binoculars. In order to provide a proper base map for the manual drawing of location and the species of the observed waterfowl, an additional UAS mission was performed 14 days before the actual ML image data capture campaign took place. This preparatory UAS mission was performed at a higher flight altitude of 120m with high overlap (90% forwardlap, 90% sidelap).

Analysis-Ready Data for Machine Learning

In order to train and run a ML model, data quality is a key issue. Several image pre-processing steps are necessary to provide analysis-ready data for the following steps of ML model development and model training. In this section, a detailed account of the pre-processing of individual UAS images is described, as well as the annotation process employed for Brenndorf. Specifically, the annotation approach utilizing a combination of non-expert laymen and annotation performed by domain experts in the field are presented. Labelbox is a professional software solution for the annotation and labeling of data for ML applications. The test site-specific taxonomy for labelling and annotation was defined by waterfowl domain experts and implemented in Labelbox for annotation.

Image Data Pre-processing

In total, 958 aerial images with a resolution of 8mm/pixel were captured by the UAS (Phantom 4 RTK) with a high overlap, which is a pre-requisite for the photogrammetric generation of high-resolution orthophotomosaics for representation and documentation of the current environmental state of the test site at Brenndorf. However, a significant drawback is that high image overlap will result in a significant amount of redundant image information. Therefore, Metashape was used to select non-overlapping images using the "*Reduce Overlap*" feature to provide a non-redundant, significantly smaller set of 171 images for more time-efficient annotation and ML classification.

Another key pre-processing step necessary to provide analysis-ready data involved the transformation of annotations into a format that is compatible with the selected ML model YOLO (see section 2.4 for more detail). Annotations in Labelbox are typically provided in JSON format, but YOLO necessitates the presence of an associated text file for each image that shares the same name and contains the class of the detected object and normalized values for the, bounding boxes' coordinates, and dimensions. The second pre-processing step involves dividing each image into smaller tiles, which is required due to the specific nature of our data. Our images have dimensions of 4864x3648 pixels, and the targeted bounding boxes are relatively small, with an average size of 92x101 pixels. As a result, the ratio of objects to the total image size is only 0.052%. This low ratio can be problematic, as YOLO typically resizes each image to 416x416 for faster training. Therefore, dividing the image into tiles serves to maintain the object's aspect ratio while resizing each tile to 416x416. Each image was tiled into 30 tiles with each tile having a size of 810x729, leading to a (target object size/input image size) ratio of 1.5%. Figure 4 demonstrates the image cropping performed.

Fig. 4

Figure 4: Demonstration of the image cropping process.
Source: own figure

Abbildung 4: *Demonstration des Bildausschnittverfahrens. Quelle: eigene Abbildung*

It is important to note that not all tiles produced during the cropping process are used for AI training. This is because, in most cases, only a small fraction of tiles actually contains waterfowl. Therefore, YOLO is only provided with the tiles that have annotated objects, and the remaining tiles are discarded.

Annotation

Annotation was performed by experienced domain experts using a detailed taxonomy of 10 classes of waterfowl and by non-expert laymen focusing on only two "high-level" classes, duck and swan. The following classes were identified by the domain experts: mallard (in German: Stockente), mute swan (Höckerschwan), "duck" (Ente), tufted duck (Reiherente), gadwall (Schnatterente), waterfowl (Wasservögel), juvenile mute swan (Höckerschwan_jung), Eurasian wigeon (Pfeifente), common teal (Krickente), and great white egret (Silberreiher).
This annotation taxonomy scheme, including the hierarchical structure of species categories (e.g., duck with mallard and common teal) and the inclusion of juvenile versions of certain species (e.g., mute swan with mute swan juvenile), makes the training process for an ML model more challenging. This is because increasing the number of waterfowl species that the machine must discriminate without increasing the dataset necessarily means giving the machine fewer training samples per class. Therefore, it is important to carefully evaluate the suitability and quality of the provided annotations for training a ML model. In some cases, the available annotations may not provide sufficient information or may introduce too much complexity, making it difficult for the model to learn effectively. Based on these considerations, it may be necessary to seek out alternative annotation schemes in order to improve the suitability of the data for training. Thus, it is not necessarily the case that the annotations are completely unsuitable for training, but

rather that their limitations and potential drawbacks need to be carefully evaluated and addressed in order to maximize their effectiveness for training of a ML model.

Machine Learning Model

The ML algorithm chosen to be used in this project is called YOLO. YOLO is a powerful state-of-the-art, real-time open source Convolutional Neural Network (CNN) object detection algorithm introduced by Redmon et al (2015).

Waterfowl knowledge base transfer

The Center for the Advancement of Spatial Informatics Research and Education (ASPIRE) at UNM collaborates with US Fish and Wildlife Service (USFWS) to develop an algorithm to automatically count and identify birds from drone imagery. Sa'Doun et al. [13] used YOLO v3 on a USFWS dataset for waterfowl detection and classification. The USFWS data volume in this US study was 13 images with 2908 unique birds. YOLO v3 was updated since 2020 to reach YOLO v5 at the time of implementing this study. Therefore, upgrading to YOLO v5 was necessary since YOLO v5 was found to have better accuracy on the same dataset [14]. The other and most important upgrade is to feed YOLO v5 with a much larger dataset than the USFWS in 2020. UNM used Zooniverse crowdsourcing service to generate more than 150 thousand new annotations to train YOLO v5 [8], [15]. We have used the weights learned by the UNM model as an initial configuration for our training. The primary benefit of this transfer learning is that it allows to train our models faster and with less data [16], while still achieving higher levels of accuracy compared to training from scratch on the Brenndorf dataset.

YOLO Training and Validation

In total, 467 objects were annotated in the Brenndorf dataset. While not considered large by ML standards, this dataset is indicative of the sample size and diversity obtainable from a single-day survey. To assess the model's performance, the dataset was divided into two subsets: a training set and a validation set. The validation set was derived as a subset of the Brenndorf data, intentionally withheld from the YOLO model. The validation images were manually selected where 17 tiles were chosen with different waterfowl composition, duck only, swan only, and mixed waterfowl. By comparing YOLO's predictions against the ground truth of the validation set, we measured the accuracy of the model using a metric called mean Average Precision at Intersection over Union (IoU 0.50) (mAP50) [17]. This metric measures how well the model detects objects in images. The 'ground truth' refers to the actual data showing where objects are located in the images, and we compare these true locations to the locations predicted by YOLO. The IoU is a measure of how much the predicted bounding box overlaps with the actual bounding box, with an IoU of 0.50 indicating a 50% overlap. Precision measures how many of the detected objects are correct, while recall measures how many of the actual objects were detected. mAP50 combines these aspects into a single number, summarizing the model's overall performance by averaging how well it balances correct detections (precision) at a 50% overlap threshold. After mAP is calculated for all objects in the validation set, a confusion matrix is created that shows the number of true positive, true negative, false positive, and false negative identifications of each waterfowl class, which correspond to commission and omission errors. Commission errors occur when the model incorrectly identifies an object (false positive), while omission errors occur when the model fails to detect an actual object (false negative).

Machine Learning Result Visualization

A key issue is to bring the results of the ML model back to the "real world," i.e. to project the YOLO detection results onto a map. This provides not only statistical, non-spatial information about the number of waterfowl detected, but also spatial information about the overall geographic distribution of waterfowl in the study area. To achieve this, a procedural framework for converting the detection output into geographic coordinates must be established and consists of two distinct steps:

1- The georeferencing of individual UAS images.
2- The projection of YOLO object detections and its estimation of real-world coordinates.

Map Projection of detected waterfowls

The automated image georeferencing pipeline relies on the following input parameters:

1- Focal length of the camera
2- Sensor size
3- Flight altitude
4- Image size
5- GNSS coordinates of the drone
6- 3D drone rotation angles (pitch, roll, yaw)

The information about the degrees of freedom can be extracted from the XML and Exif metadata that are attached to an image. The metadata will then be incorporated to calculate a transformation matrix to translate pixel locations to GNSS coordinates. Two Python libraries were used: CameraTransform and GDAL. CameraTransform was used to build a camera model to calculate the transformation matrix from pixel to GNSS, whereas GDAL was used to apply the transformation to the image to generate a raster layer in projected map coordinates. Figure 5 shows the principle of image georeferencing and Figure 6 shows our results from this process.

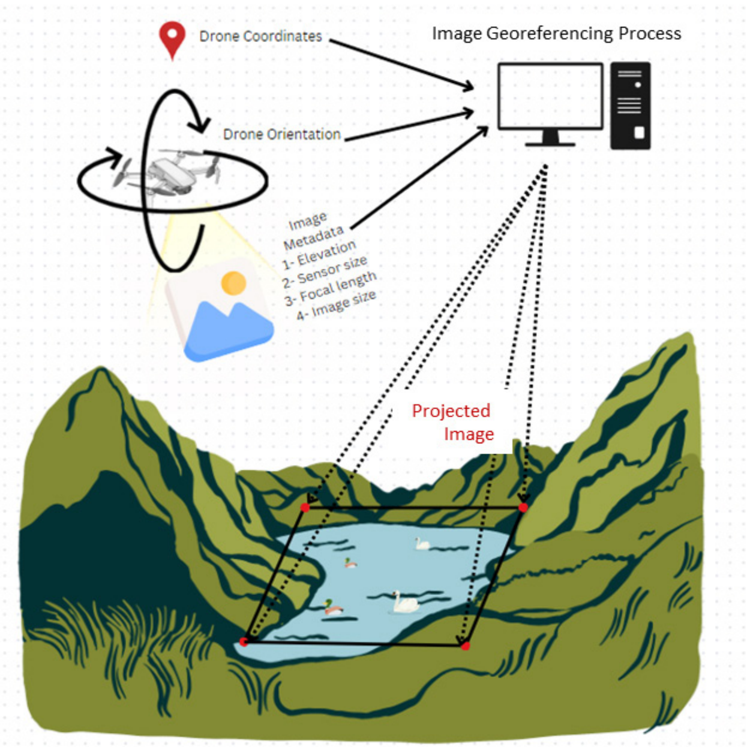

Fig. 5

Figure 5: Image georeferencing process.
Source: own figure

Abbildung 5: *Prozess der Georeferenzierung von Bildern.*
Quelle: eigene Abbildung

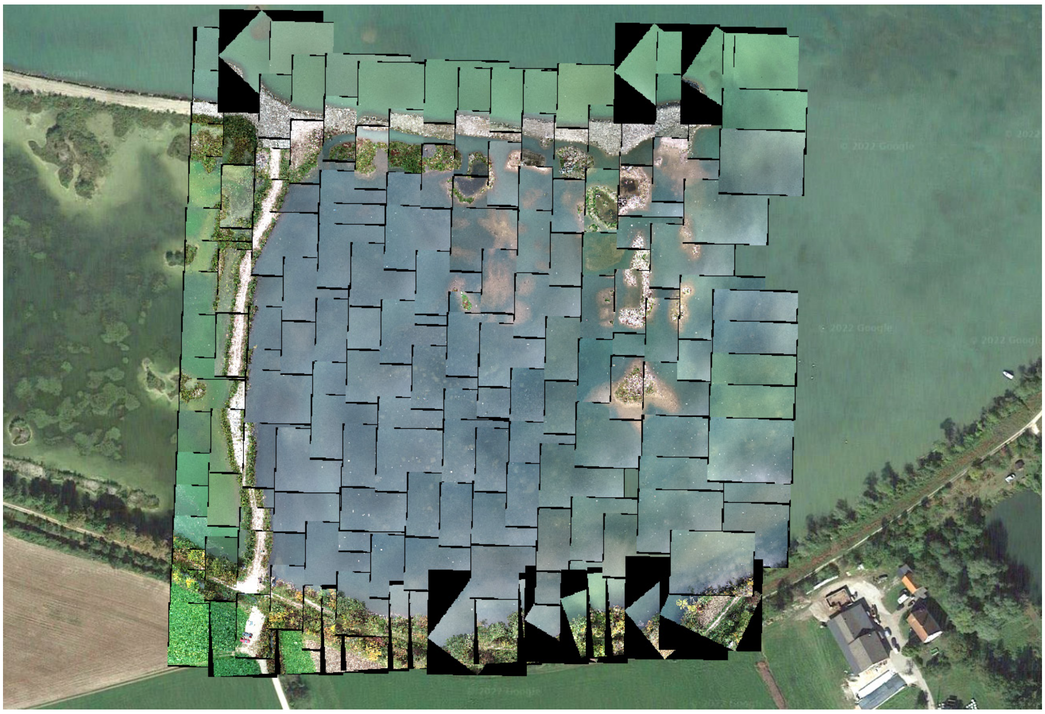

Fig. 6

Figure 6: Georeferenced and projected individual UAS image tiles. Source: own figure

Abbildung 6: *Georeferenzierte und projizierte einzelne UAS-Bildkacheln. Quelle: eigene Abbildung*

YOLO detections come in a normalized format. For each image where YOLO detected objects, a standardized text file containing detection information was produced. Two conversions are required here for georefercing the detection results:

1- Convert YOLO's normalized detections of tiles to original image pixel coordinates,
2- Convert pixel locations in the original image to GNSS coordinates (point vector).

Table 2 represents the vector points of detection sorted in the attribute table. Based on the tile name and the size of the tile image it is possible to infer the specific section of the original image to which the tile corresponds by examining its suffix. A reverse calculation is then performed and the pixel locations in the original image are retrieved. Figure 7 demonstrates normalized tile detection to original image pixel conversion.

Tab. 2

Class	X	Y	Confidence (%)	Latitude	Longitude	Original Image
swan	2702	3371	92.3	46.6350604	14.6000256	100_0627_0438.JPG
duck	4657	2728	82.1	46.6356269	14.6003050	100_0627_0448.JPG
duck	2203	282	79.9	46.6354206	14.6002799	100_0628_0002.JPG

Table 2: Example of attribute table for the generated vector point shapefile

Tabelle 2: *Beispiel einer Attributtabelle für das erzeugte Vektorpunkt-Shapefile*

Then, the transformation model from CameraTransform is used to calculate geographic coordinates. GeoPandas [18] is then used to generate vector points from the derived coordinates.

By completing these steps, YOLO detections can be mapped, allowing further visualization and analysis of their spatial distribution. The spatially explicit visualization of automated waterfowl detections provides, in addition to plain numbers and non-spatial statistics, other important insights into distribution patterns, potential species interactions and behaviors, as well as qualitative and quantitative comparisons with traditional field survey observations.

```
tile: 100_0627_0386_02_03 (810x729)
suffix: 02_03 (vertical position, horizontal position)

class_code   Xcenter   Ycenter   Width    Height
     1       0.083951 0.506859  0.08321 0.059085
                          ↓
Original image: 100_0627_0386 (4864x3648)

class_code   Xcenter   Ycenter   Width    Height
     1         1688      1099      67        43
```

Fig. 7

Figure 7: Normalized tile detections to original image pixel conversion

Abbildung 7: *Umrechnung von normalisierten Bildkacheln in Originalbildpixel*

RESULTS

Our results include the observed and mapped locations of waterfowl from the water's edge identified by a domain expert performing a traditional field survey, the annotation process, the ML classification, and finally the spatial map-based visualization of the automatically identified waterfowl.

Traditional waterfowl survey – Field mapping

Figure 8 shows the results of field mapping with hand-drawn point locations on a provided map of the observed waterfowl species at Brenndorf. Table 3 gives a summary of the observed waterfowl species at Brenndorf. In total, 209 individual waterfowl were detected by one experienced expert in the context of the traditional ground-based field survey.

Fig. 8

Figure 8: Example of hand-drawn locations of observed waterfowl species from traditional field mapping by an experienced domain expert at Brenndorf. Field observation was performed within the same time frame as UAS image capture. Source: own figure

Abbildung 8: *Beispiel für handgezeichnete Standorte der beobachteten Wasservogelarten aus der traditionellen Feldkartierung durch einen erfahrenen Fachmann in Brenndorf. Die Feldbeobachtung wurde im gleichen Zeitrahmen wie die UAS-Bilderfassung durchgeführt. Quelle: eigene Abbildung*

Tab. 3

	Common name		
Latin	**English**	**German**	**Count**
Anas platyrhynchos	mallard	Stockente	51
Cygnus olor	mute swan	Höckerschwan	47
Mareca strepera	gadwall	Schnatterente	42
Aythya fuligula	tufted duck	Reiherente	35
Anas crecca	common teal	Krickente	8
Calidris pugnax	ruff	Kampfläufer	8
Podiceps cristatus	great crested grebe	Haubentaucher	4
Phalacrocorax carbo	great cormorant	Kormoran	3
Larus michahellis	yellow-legged gull	Mittelmeermöwe	2
Gallinago gallinago	common snipe	Bekassine	2
Mareca penelope	Eurasian wigeon	Pfeifente	2
Mergus merganser	goosander	Gänsesäger	1
Ardea alba	great white egret	Silberreiher	1
Ardea cinerea	grey heron	Graureiher	1
Aythya ferina	common pochard	Tafelente	1
Anas acuta	northern pintail	Spießente	1
	Total count		209

Table 3: Summary of the observed waterfowl species distribution as result of the expert field mapping at Brenndorf

Tabelle 3: *Zusammenfassung der beobachteten Wasservogelartenverteilung als Ergebnis der Experten-Feldkartierung in Brenndorf*

Annotation

The final dataset for annotation consisted of 171 non-overlapping images containing waterfowl. A total of 434 waterfowl were annotated by the domain experts. The classes duck and waterfowl were introduced for complex cases where a clear species identification was not possible. The introduction of the duck and waterfowl classes by domain experts highlights the challenges associated with accurate species-level identification in waterfowl imagery. This practical approach ensured comprehensive data collection while acknowledging the limitations of current identification methods. The decision to review ambiguous cases further emphasizes the commitment to data quality and the potential for future expert refinement. The overall distribution of domain expert annotation results is shown in Table 4.

Tab. 4

	Common name			
Latin	**English**	**German**	**Count**	**Share (%)**
Anas platyrhynchos	mallard	Stockente	156	35.9
Cygnus olor	mute swan	Höckerschwan	101	23.3
Anatinae	duck*	Ente*	60	13.8
Aythya fuligula	tufted duck	Reiherente	52	12.0
Mareca strepera	gadwall	Schnatterente	34	7.8
-	waterfowl*	Wasservogel*	12	2.8
Cygnus olor	mute swan juvenile	Höckerschwan Jung	10	2.3
Mareca penelope	Eurasian wigeon	Pfeifente	7	1.6
Anas crecca	common teal	Krickente	1	0.2
Ardea alba	great white egret	Silberreiher	1	0.2
			434	100.0

Table 4: Domain expert taxonomy and annotation distribution. The class duck serves as an overview class where a detailed duck species classification can not be achieved. The class waterfowl represent a detection, but no further classification was possible for the annotator

Tabelle 4: *Taxonomie eines Experten und Verteilung der Anmerkungen. Die Klasse Ente dient als Übersichtsklasse, wenn eine detaillierte Klassifizierung der Entenarten nicht möglich ist. Die Klasse Wasservogel stellt einen Nachweis dar, aber eine weitere Klassifizierung war für den Kommentator nicht möglich*

The same Brenndorf dataset was also annotated by non-expert laymen with the goal to annotate at the species level, i.e. non-expert laymen were able to distinguish between two classes of waterfowl: duck and swan. In total, 467 waterfowl were annotated by laymen, indicating that they were able to count and identify the same amount of waterfowl as domain experts. However, it should be noted that the laymen were not able to automatically distinguish between all the species classes as effectively as the domain experts. Non-expert laymen annotation distribution is shown in Table 5.

Tab. 5

Species	Count	Share (%)
duck	353	75.6
swan	114	24.4
	467	100.0

Table 5: Non-expert laymen "high-level" taxonomy of waterfowl and their distribution

Tabelle 5: *„High-Level"-Taxonomie der Wasservögel und ihre Verteilung durch Laien*

Machine Learning Classification

We chose a 7% training/validation split to maximize the training data available for model development. The Brenndorf dataset for ML classification comprises 5130 tiles (171 x 30), out of which 242 tiles contained bird annotations. Thus, 17 tiles were allocated for validation. Within these tiles, a total of 38 waterfowl were present, including 10 swans out of 114 and 28 ducks out of 353.

This approach ensured that the model is robust and effective in projecting waterfowl locations on a map. A larger validation set would have reduced the training data, potentially compromising model accuracy. As the size of our dataset grows in the future, the validation set size can be increased for more robust evaluation without compromising training

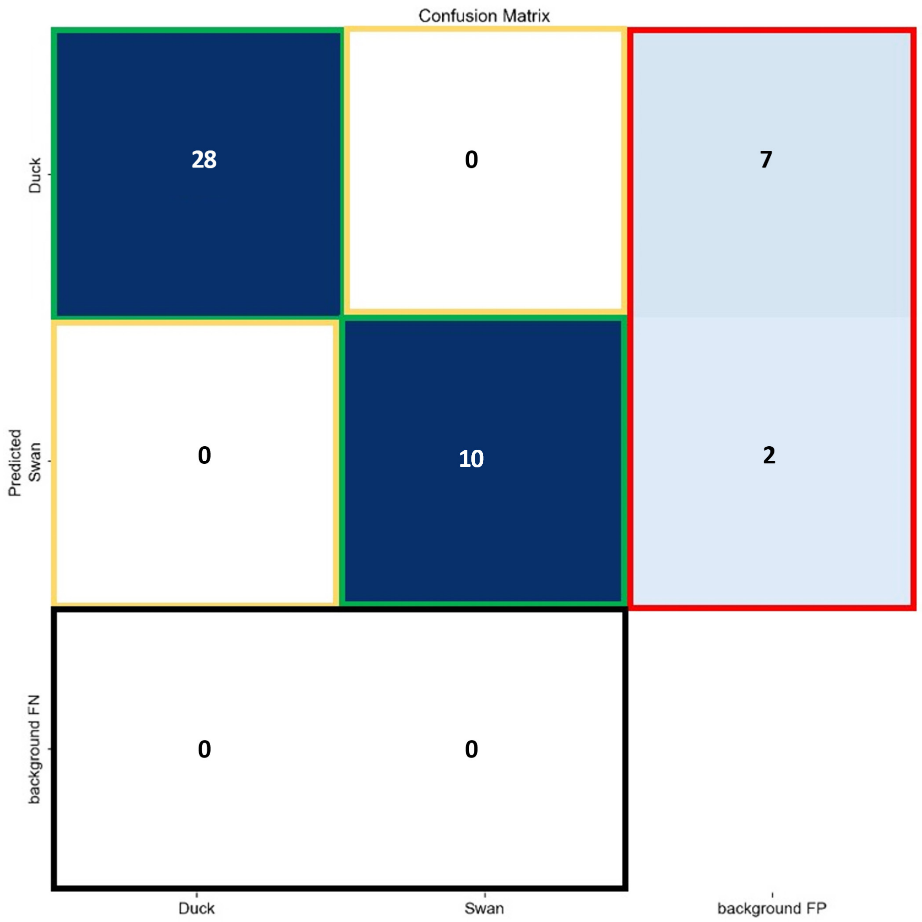

Fig. 9

Figure 9: Confusion matrix representing validation results. Green: True positive (correct detection and classification). Yellow: True negative (correct detection but wrong classification). Black: Undetected. Red : Background detected as waterfowl. Source: own figure

Abbildung 9: *Konfusionsmatrix zur Darstellung der Validierungsergebnisse. Grün: Wahr positiv (korrekte Erkennung und Klassifizierung). Gelb: Wahr negativ (korrekte Erkennung, aber falsche Klassifizierung). Schwarz: Unerkannt. Rot: Hintergrund als Wasservogel erkannt. Quelle: eigene Abbildung*

effectiveness. Throughout the training process, 20 epochs were conducted. The results are presented in Figure 9, which exhibits the confusion matrix obtained from the model evaluation. The confusion matrix revealed that all ducks and swans in the validation set were accurately detected and classified (28/28 ducks and 10/10 swans). Nonetheless, the model exhibited a tendency to misclassify background elements as waterfowl, resulting in 7 false detections of ducks (25%) and 2 false detections of swans (20%). Consequently, the accuracy of the model on the validation data set can be calculated as 28/35 = 80% for the class duck and 10/12 = 83.3% for the class swan. Although the size of the validation set is relatively small, it is important to note that it is currently sufficient to complete the projection pipeline. However, it is acknowledged that a larger validation set would be more beneficial. In the future, the incorporation of a significantly larger volume of data may further improve the accuracy and reliability of the model.
Figure 10 shows samples of the classification results demonstrating perfect detection and classification for each class as well as background false classification.

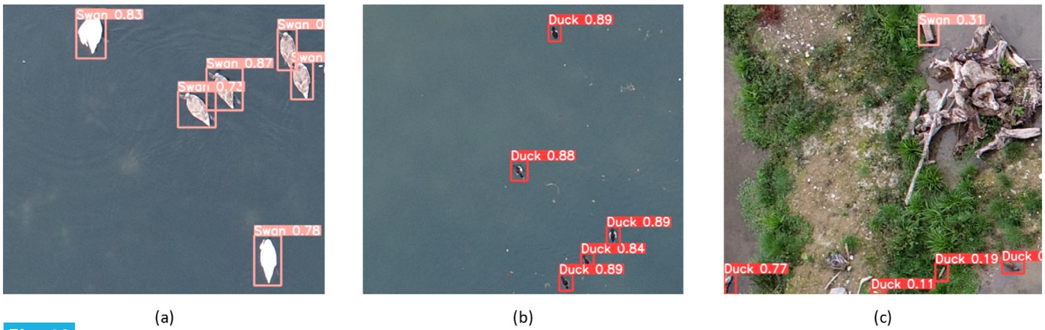

Fig. 10

Figure 10: Correct detection and classification of class swan (a); Correct detection and classification of class duck (b); Background objects misclassified as ducks or swans (c). Source: own figure

***Abbildung 10:** Korrekte Erkennung und Klassifizierung der Klasse Schwan (a); Korrekte Erkennung und Klassifizierung der Klasse Ente (b); Hintergrundobjekte, die fälschlicherweise als Enten oder Schwäne klassifiziert wurden (c). Quelle: eigene Abbildung*

Map-based Visualization

Figure 11 shows the spatial distribution of waterfowl that were annotated on individual image tiles in Labelbox. Such annotation data can be used not only for ML training and validation purposes, but also for collaborative mapping of waterfowl based on high-resolution UAS imagery.

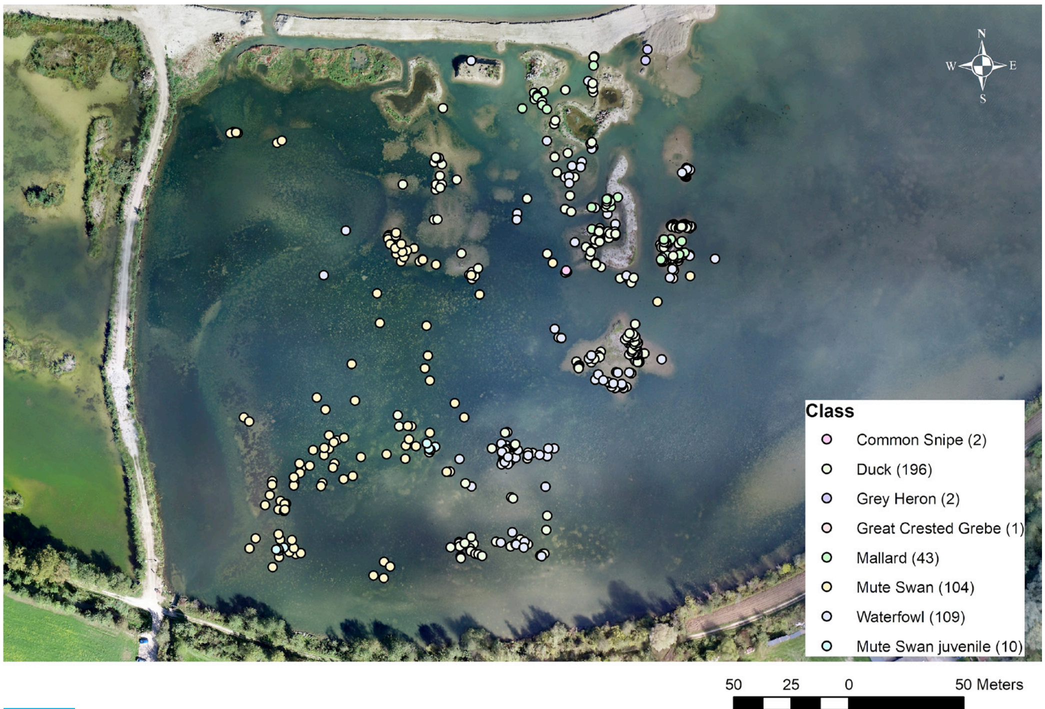

Fig. 11

Figure 11: Spatial distribution of labels annotated in Labelbox. Source: own figure

***Abbildung 11:** Räumliche Verteilung der in der Labelbox kommentierten Labels. Quelle: eigene Abbildung*

Figure 12 shows the YOLO detection results and spatial distribution of the two high-level waterfowl classes duck and swan in the project area in the late morning of October 6, 2022. In addition, the current environmental conditions of the habitat in terms of vegetation phenology, vegetation structure, riparian morphology, and water conditions are documented by a high-resolution orthophotomosaic that serves as a base map.

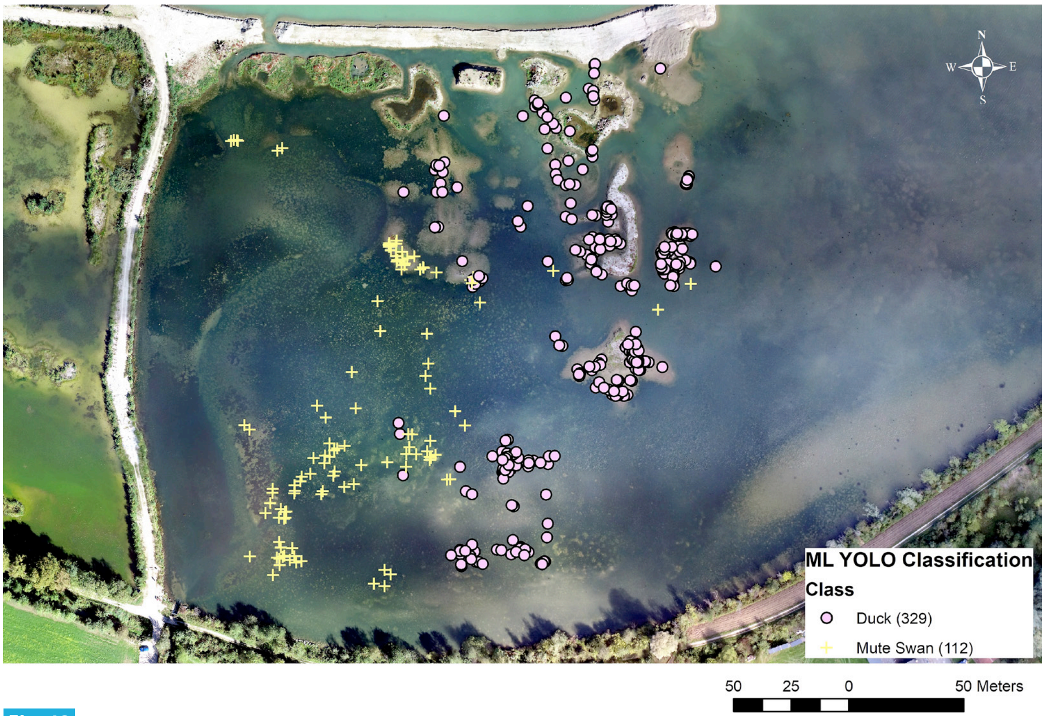

Fig. 12

Figure 12: YOLO detection results at Brenndorf. Source: own figure

Abbildung 12: *YOLO-Erkennungsergebnisse in Brenndorf. Quelle: eigene Abbildung*

SUMMARY AND DISCUSSION

In this study, we presented a framework for non-invasive automated waterfowl detection using the state-of-the-art machine learning CNN YOLO and high-resolution UAS imagery. The pipeline involved several stages, each contributing to the overall effectiveness of the process. The pipeline developed here serves as a proof of principle rather than a fully automated and ready-to-use application. Data acquisition was the initial stage of the pipeline, where we utilized UAS to capture high-resolution images for waterfowl detection. By conducting UAS missions at a flight altitude of 30m in the Brenndorf test site, we obtained a non-invasive dataset that formed the basis for subsequent stages. Following data acquisition, the collected images underwent data pre-processing to optimize their suitability for training and validation. We employed Metashape software to reduce overlap within the image set and a tiling scheme to increase the object-to-image ratio. Image data were annotated using Labelbox and converted into text file format for YOLO analysis, preparing it for the training stage. The AI training and validation stage incorporated the use of YOLOv5 and a transfer learning approach by leveraging knowledge from a previous project at UNM that utilized YOLOv5 for waterfowl detection and classification on a large dataset. This approach yielded promising results on our small dataset, with the model achieving high classification accuracy of 80% and 83% for the high-level classes duck and mute swan, with all errors relating to commission caused by false detection of landscape objects (e.g., grass clumps) as waterfowl.

Table 6 provides a summary of the overall results of this pilot project. The automated machine learning-based detection and classification with YOLO shows highly promising results for two "high level" classes with high detection and classification accuracies.

Overall accuracy decreased dramatically when considering a more complex taxonomy, which can be explained by the lack of a sufficient amount of training data, especially for rare species.

Tab. 6

Test Site Brenndorf 10/6/2022	Number of ML classes	Number of labels	Detection Accuracy (%)	Classification Accuracy (%)
Non-expert Laymen	2	467	94.4	85
Domain Experts	10	434	24.9	14.7
Test Site Brenndorf 10/6/2022	**Number of field classes**	**Number of field observations**		
Expert Field Survey	16	209		

Table 6: Comparison of field observations versus ML detection and classification accuracy. Blue: Annotated labels by non-expert laymen and domain experts and the related ML results; Green: Results achieved by expert field survey

Tabelle 6: *Vergleich der Feldbeobachtungen mit der ML-Erkennungs- und Klassifizierungs-genauigkeit. Blau: Von Laien und Fachleuten kommentierte Beschriftungen und die entsprechenden ML-Ergebnisse; Grün: Ergebnisse der Feldbeobachtung durch Experten*

More than twice as many individual waterbirds were identified based on the high-resolution UAS imagery collected concurrently with the traditional ground-based expert field survey. It is important to critically reflect upon and understand the advantages and challenges of both the traditional field survey and the proposed ML-based framework. We identify six key points: (1) Field observers and UAS operate with different viewing perspectives and angles. These are determined for the field observer by body height, which provides a predominant horizontal view of waterfowl and their characteristic features. UAS provide a vertical perspective of waterfowl as images with very high resolution of several millimeters, depending on the camera sensor and flight altitude. These different viewing angles and perspectives can be seen as complementary to identify waterfowl species characteristics. In the annotation process, the availability of the vertical view on waterfowl are another challenge for domain experts, as they are currently used to identify waterfowl characteristics only from the horizontal side view. (2) UAS provide an excellent overview of a project area, whereas field observers are clearly limited by accessibility as well as by the size, morphological characteristics, and vegetation structure of the shoreline of the water body. The larger and more complex the project area is, the more significant is the advantage of the UAS approach to cover large areas, large parts of which are inaccessible from the ground. (3) For annotation and ML classification, a project-area-specific waterfowl taxonomy was defined in close collaboration with domain experts. Non-experts are able to recognize the same amount and number of easily identifiable waterfowl as domain experts, such as “duck” and “swan,” as well as “waterfowl” in general. However, only experienced domain experts are able to annotate waterfowl at the subspecies level. The digital annotation process and workflow also allow any waterfowl that could not be identified by non-experts to be forwarded to a domain expert. With this approach, the general distribution of “waterfowl” can be documented in a first step, with a refinement of the species classification by experts in a second follow-up procedure. (4) The quality of the ML classification depends strongly on the amount and quality of the training data. Here, the level of species taxonomy is important to consider, because for each level of species taxonomy, a sufficient amount of training data must be available. It can be clearly seen that the performance of the ML model decreases significantly as the number and level of species hierarchy increases. (5) After completing the training and validation stage, the pipeline moved to the map projection stage. Here, we constructed a projection model and pipeline to visualize the YOLOv5 detections on a map. This spatial representation of the detected birds closes the loop and spatially represents the ML results on a map. Thus, not only the pure number of observed waterbird classes in tabular form, which is a typical result of traditional surveys, but also where these classes are observed, are some of the most promising results of this approach for a comprehensive understanding of the spatial distribution and spatial patterns of waterfowl and their geographic context to the habitat. Finally, (6) digital UAS data such as the mission plan and the captured images represent a geo-referenced spatiotemporal representation and documentation of the project area with an explicit time stamp of the survey performed. Such digital data can now be further used to build a digital archive for quality assurance and documentation and thus contribute as a thematic building block to a more comprehensive spatial data infrastructure for environmental monitoring of waterfowl and related habitats.

CONCLUSION AND FUTURE WORK

The results of this study demonstrated the potential effectiveness of the proposed pipeline for automated, non-invasive waterfowl detection from UAS imagery. By integrating data acquisition, pre-processing, AI training and validation, and projection, we made progress toward accurate and reliable waterfowl detection and classification. Despite limitations such as the controlled environment of the UAS mission and the focus on two specific classes of waterfowl, the study provides valuable insights for improved wildlife management, conservation, and habitat mapping by taking advantage of new digital technologies.

This framework may be expanded in the future to detect and classify additional waterfowl species or multiple classes simultaneously. Expanding the framework requires careful consideration of several factors. First, it is important to prioritize target species based on ecological relevance, conservation status, and data availability. Second, developing robust object detection and classification models for new species requires the collection and annotation of high-quality training data. In addition, training data augmentation offers new ways to increase the amount of training data. This involves the automatic modification of captured images - such as rotating, adjusting brightness, adding noise, etc. - to create a larger and more diverse set of training data. Third, exploring transfer learning techniques can help leverage knowledge from existing models to accelerate development of models for new species. Spatial distribution data generated by Labelbox provide a basis for in-depth ecological analysis and conservation planning. By combining this information with other environmental factors captured in the UAS imagery (e.g., habitat type, water bodies, land use and land cover), it is possible to identify important waterfowl habitat, migration patterns, and potential threats. In addition, the ability to conduct collaborative mapping helps build a community-driven approach to conservation and enables better monitoring of waterfowl populations.

The discrepancy between expert and non-expert annotations underscores the challenges inherent in waterfowl monitoring projects, especially in complex areas such as species identification. While non-expert laypersons were able to differentiate between ducks and swans, their limited taxonomic knowledge likely hindered their ability to accurately classify other waterfowl species or to further classify duck or swan species. This suggests that while non-expert annotation can be a valuable tool for data collection, careful quality control measures and expert validation are essential to ensure data reliability and for future scaling and expansion of such a methodology.

In addition, evaluating the performance of the proposed pipeline in different environments and under different conditions would increase its robustness and generalizability. High-resolution UAS imagery also adds value by digitally documenting the current state of the environment at the time of the flight. Researchers and volunteers can more accurately count birds from the imagery and refer back to the imagery if questions arise later. The images also provide a wealth of contextual information (e.g., vegetation condition/type, water levels, climatic conditions, etc.) that would be difficult to replicate with field surveys. Last but not least, close collaboration between domain experts, UAS and ML researchers throughout the project is a key success factor, starting with user-centered and domain-specific requirements analysis, through training and annotation, to results evaluation and validation [19].

The presented framework provides a foundation to support traditional waterbird surveys with new and innovative methods and provides promising challenges for further interdisciplinary research and development of automated waterbird detection systems using UAS imagery in Carinthia and beyond.

ACKNOWLEDGEMENTS

This study has been financed by the Regional Government of Carinthia, Department 8 – Environment, Energy and Nature Protection.

REFERENCES

[1] S. J. Dundas, M. Vardanega, P. O'Brien, and S. R. McLeod, "Quantifying waterfowl numbers: comparison of drone and ground-based survey methods for surveying waterfowl on artificial waterbodies," *Drones*, vol. 5, no. 1, p. 5, Jan. 2021, doi: 10.3390/drones5010005

[2] S. Wagner and W. Petutschnig, "Wasservogelzählung in Kärnten 2021," *Carinthia II*, vol. : 211_131_1 pp. 177-186, 2021.

[3] IWC, "About the International Waterbird Census." Wetlands.org, Accessed Aug. 26, 2024. [Online]. Available: https://iwc.wetlands.org/index.php/aboutiwc

[4] D. Wen, L. Su, Y. Hu, Z. Xiong, M. Liu, and Y. Long, "Surveys of large waterfowl and their habitats using an unmanned aerial vehicle: a case study on the Siberian crane," *Drones*, vol. 5, no. 4, p. 102, Sep. 2021, doi: 10.3390/drones5040102

[5] S.-J. Hong, Y. Han, S.-Y. Kim, A.-Y. Lee, and G. Kim, "Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery," *Sensors*, vol. 19, no. 7, p. 1651, Apr. 6 2019, doi: 10.3390/s19071651

[6] Z. Tang, et al., "sUAS and Machine Learning Integration in Waterfowl Population Surveys," *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov. 2021, doi: 10.1109/ictai52525.2021.00084

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016, doi: 10.1109/cvpr.2016.91

[8] R. L. Converse, C. D. Lippitt, S. E. Sesnie, G. M. Butler and D. R. Stewart, "Observer variability in manual-visual interpretation of UAS imagery of wildlife, with insights for deep learning applications," unpublished.

[9] R. L. Converse, "Drones for ducks," www.zooniverse.org. Accessed Aug. 26, 2024. [Online]. Available: https://www.zooniverse.org/projects/rowan-aspire/drones-for-ducks

[10] European Union, 2019/945, Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02019R0945-20200809

[11] European Union, 2019/947, Commission Implementing Regulation (EU) 2019/947 of 24 May 2019 on the rules and procedures for the operation of unmanned aircraft. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02019R0947-20220404

[12] Rechtsinformationssystem des Bundes (RIS), 21/2019, Kundmachung der Landesregie-rung vom 26. März 2019, Zl. 01-VD-LG-1883/1-2019, über die Wiederverlautbarung des Kärntner Nationalpark- und Biosphärenparkgesetzes. [Online]. Available: https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=LrK&Gesetzesnummer=20000339&FassungVom=2024-08-18

[13] M. M. Sa'doun, C. D. Lippitt, G. Paulus, and K. Anders, "A comparison of convolutional neural network architectures for automated detection and identification of waterfowl in complex environments," *GI_Forum*, vol. 1, pp. 152–166, 2021, doi: 10.1553/giscience2021_02_s152

[14] U. Nepal and H. Eslamiat, “Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs,” *Sensors*, vol. 22, no. 2, p. 464, Jan. 8 2022, doi: 10.3390/s22020464

[15] R. L. Converse, C. D. Lippitt, S. E. Sesnie, G. M. Butler and D. R. Stewart, “Optimizing deep learning training data for automating aerial wildlife monitoring,” unpublished

[16] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J. Big Data*, vol. 3, no. 1, p. 9, May 2016, doi: 10.1186/s40537-016-0043-6

[17] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, “UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios,” *Sensors*, vol. 23, no. 16, pp. 7190–7190, Aug. 15 2023., doi: 10.3390/s23167190

[18] J. Van den Bossche, Jul. 02, 2024, “geopandas/geopandas: v1.0.1,”, distributed by Zenodo, https://zenodo.org/records/12625316, doi: 10.5281/zenodo.12625316

[19] C. D. Lippitt, D. A. Stow, and K. C. Clarke, “On the nature of models for time-sensitive remote sensing,” *Int. J. Remote Sens.*, vol. 35, no. 18, pp. 6815–6841, Sep. 2014, doi: 10.1080/01431161.2014.965287

ABOUT THE AUTHORS

Gernot Paulus
Spatial Informatics for ENvironmental Applications (SIENA), Carinthia University of Applied Sciences, Villach, Austria
E-Mail: g.paulus@fh-kaernten.at

Mohammed Sa’Doun
SIENA, Carinthia University of Applied Sciences, Villach, Austria

Karl-Heinrich Anders
SIENA, Carinthia University of Applied Sciences, Villach, Austria

Ulf Scherling
SIENA, Carinthia University of Applied Sciences, Villach, Austria

Werner Petutschnig
Office of the Carinthian Provincial Government, Dept. 8 - Environment, Energy and Nature Conservation, Klagenfurt, Austria

Johann Wagner
Office of the Carinthian Provincial Government, Dept. 8 - Environment, Energy and Nature Conservation, Klagenfurt, Austria

Christopher D. Lippitt
CMS-RS, University of New Mexico Department of Geography and Environmental Studies, Albuquerque NM, 87131, USA

BACK TO CONTENT