

# Silbermöwe oder Heringsmöwe? - Klassifikation von Seevögeln in UAS-Aufnahmen

KATRIN GÖLZ<sup>1</sup>, SYLVIA HOCHSTUHL<sup>1,2</sup> & ANTJE THIELE<sup>1,2</sup>

*Zusammenfassung: Für eine effiziente Erfassung der Vogelbestände, werden in dieser Arbeit lernbasierte Verfahren für die Klassifikation unterschiedlicher Seevögel in Luftbilddaufnahmen getestet und verglichen. Ausgewertet werden zwei Datensätze, welche über Spiekeroog und Baltrum mittels unbemannter Flugsysteme aufgenommen wurden. Für die Unterteilung von Bildausschnitten in die Klassen Seevogel und Hintergrund kommen ein SVM-basierter und ein CNN-basierter Klassifikator zum Einsatz. Beide Ansätze liefern mit Gesamtgenauigkeiten von über 92% sehr gute Ergebnisse, wobei die Klassifikationsleistung des CNNs die des SVM-basierten Verfahrens leicht übersteigt. Die CNN-basierte Klassifikation wird darüber hinaus für die Unterscheidung von bis zu sechs unterschiedlichen Vogelarten eingesetzt. Während Vogelarten, für welche eine ausreichende Anzahl an Trainingsbeispielen verfügbar ist, mit recall-Werten von bis zu 96% erkannt werden, liegen die recall-Werte der im Training unterrepräsentierten Klassen deutlich niedriger, sodass balanced accuracy-Werte von bis zu 62% und 84% erzielt werden.*

## 1 Einleitung

Mehr als 200 Vogelarten sind als Brutvögel in Niedersachsen heimisch. Eine kontinuierliche Erfassung der Daten zu Vorkommen, Verbreitung und Bestandsentwicklung der Brutvögel ist für die Naturschutzarbeit von Behörden und Verbänden essentiell. So dienen die erhobenen Daten als wichtige Entscheidungsgrundlagen bei der Ausweisung von Schutzgebieten, deren Abgrenzung und Zonierung sowie der Planung von Schutzmaßnahmen. Neben der herkömmlichen und aufwändigen Methode, Vögel bei Feldbegehungen zu zählen, werden zunehmend hochaufgelöste Luftbilddaufnahmen für die Vogelzählung herangezogen, welche mittels unbemannter Flugsystemen (engl. Unmanned Aircraft System (UAS)) aufgenommen werden. Der Einsatz von UAS-Systemen bietet den Vorteil der flächendeckenden, effizienten und kostengünstigen Datenerfassung, durch welche die Brutvögel laut SARDÀ-PALOMERA et al. (2012) und GRENZDÖRFFER (2013) nicht gestört werden. Für die Vogelzählung werden die Luftbilddaufnahmen in der Regel von fachkundigem Personal manuell ausgewertet. Da in großen Brutgebieten teilweise mehrere Tausend Vögel brüten, ist die manuelle Detektion und Klassifikation aller Vogelarten allerdings eine langwierige Aufgabe. Die automatische Bilddatenauswertung hat somit das Potential den Arbeitsaufwand der Vogelzählung deutlich zu reduzieren.

Der erfolgreiche Einsatz automatischer Detektionsverfahren für das Zählen von Vögeln in Luftbilddaufnahmen wird in unterschiedlichen Studien demonstriert (GRENZDÖRFFER 2013; DIAZ-DELGADO et al. 2017; AFÁN et al. 2018; HODGSON et al. 2018; RUSH et al. 2018). Dabei werden

---

<sup>1</sup> Karlsruher Institut für Technologie, Institut für Photogrammetrie und Fernerkundung, Englerstr. 7, 76131 Karlsruhe, E-Mail: Katrin.Goelz@gmx.de

<sup>2</sup> Fraunhofer IOSB, Ettlingen, Gutleuthausstr. 1, 76275 Ettlingen, E-Mail: [Sylvia.Hochstuhl, Antje.Thiele]@iosb.fraunhofer.de

bereits teilweise Methoden aus dem Bereich des maschinellen Lernens eingesetzt. So werden in der Arbeit von DIAZ-DELGADO et al. (2017) Möwen in UAS-Bildaufnahmen mittels Support Vector Machine (SVM) und Random Forest detektiert und automatisch gezählt. Die so erzielten Ergebnisse werden anhand von Bodenbeobachtungen validiert und ergeben für beide Ansätze hohe Gesamtgenauigkeiten von 82 % bis 98 %. Damit ist die automatische Klassifikation nur wenig ungenauer als die zeitaufwändige manuelle Auswertung der UAS-Bildaufnahmen. Die Random Forest-basierte Klassifikation wird auch von AFÁN et al. (2018) für die Erkennung des Braunen Sichlers verwendet. In diesem Anwendungsfall ist die Gesamtgenauigkeit mit 46% jedoch unter anderem aufgrund des dunklen Gefieders der Vögel, niedrig. Eine weitere vielversprechende Methode für die automatische Vogelzählung ist die CNN (Convolutional Neural Network)-basierte Klassifikation. So stellen HONG et al. (2019) eine Studie vor, in welcher Vögel in UAS-Aufnahmen mittels der Detektionsnetze Yolo, Faster Region-based CNN, Region-based Fully Convolutional Network, Single Shot MultiBox Detector und Retinanet mit durchschnittlichen Genauigkeiten (average precision) zwischen 85% und 95% erkannt werden. Auch in der Arbeit von KELLENBERGER et al. (2021) bringt die CNN-basierte Detektion von Seevögeln gute Ergebnisse, die durch die Einbeziehung von Vorwissen über die räumliche Verteilung von Vogelkolonien mittels Markov Random Fields noch verbessert werden können.

Für die Detektion von Seevögeln in Brutgebieten auf den ostfriesischen Inseln kommen in dieser Arbeit die SVM-basierte und CNN-basierte Klassifikation zum Einsatz. In den betrachteten Brutgebieten nisten verschiedene Arten von Seevögeln. Um Rückschlüsse auf die Populationsentwicklung der unterschiedlichen Vogelarten zu ziehen, wird in dieser Arbeit neben der reinen Erkennung von Vögeln, die in den angeführten Studien schwerpunktmäßig thematisiert wird, ein weiterer Fokus auf die Klassifikation der Seevogelarten gelegt. Eine spezielle Herausforderung dabei ist die unausgewogene Klassenverteilung. Während für einige Vogelarten mehrere Tausend Trainingsbeispiele zur Verfügung stehen, sind andere Arten nur schwach repräsentiert.

Die vorliegende Arbeit ist folgendermaßen gegliedert: Die Datenerhebung und -aufbereitung der UAS-Aufnahmen wird in Abschnitt 2 vorgestellt. In Abschnitt 3 folgt die Beschreibung der verwendeten Methodik für die automatische Klassifikation. Die erzielten Ergebnisse werden in Abschnitt 4 vorgestellt und diskutiert. Abschließend werden die Kernaussagen in Abschnitt 5 zusammengefasst und ein Ausblick auf zukünftige Arbeiten wird gegeben.

## 2 Datengrundlage

Als Datengrundlage dienen in dieser Arbeit zwei Datensätze des Niedersächsischen Landesbetriebs für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN), welche 2018 und 2021 über zwei verschiedenen Untersuchungsgebieten auf Spiekeroog und Baltrum mittels UAS aufgenommen wurden. Die beflogenen Gebiete sind in Abb. 1 dargestellt. Im Folgenden wird das Vorgehen der Datenerfassung und der visuellen Auswertung der Bilder für die Generierung von Trainingsdaten vorgestellt.

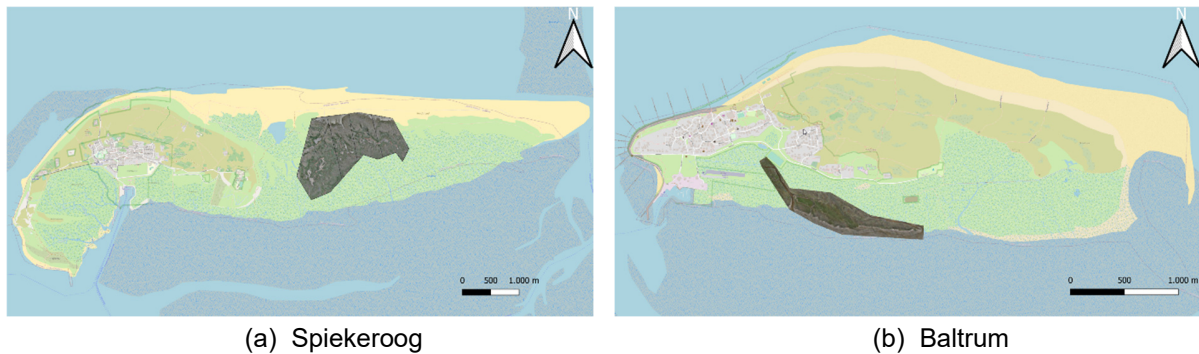


Abb. 1: Aufgenommene Brutgebiete auf Spiekeroog und Baltrum

## 2.1 Datenerfassung

Im Rahmen der UAS-Messkampagnen, deren Ziel insbesondere die Zählung brütender Seevögel ist, wurden die Untersuchungsgebiete jeweils zweimal an einem Tag innerhalb des Brutzeitraums (Mitte April bis Anfang Juli) befliegen. Die zeitlich leicht versetzte Aufnahme des gleichen Gebietes soll dazu dienen brütende von nichtbrütenden Seevögeln zu unterscheiden. Die Befliegungen wurden um die Mittagszeit durchgeführt, um einen möglichst kurzen Schattenwurf der beobachteten Vögel zu gewährleisten. Optimalerweise sollte der Himmel während der Befliegung bewölkt sein, sodass diffuse Lichtverhältnisse vorherrschen. Angaben zu dem verwendeten System für die Bildaufnahme sind Tab. 1 zu entnehmen. Um die Generierung von Orthofotos, welche das beobachtete Gebiet maßstabsgetreu und verzerrungsfrei darstellen, zu ermöglichen, wurden die Gebiete mit einer Bildüberlappung von etwa 80% längs und 60% quer zu der Flugrichtung aufgenommen. Über den GNSS-Empfänger, welcher auf dem UAS angebracht ist, wird die Position zu jedem Aufnahmezeitpunkt ermittelt und mittels Real Time Kinematic (RTK) korrigiert. Damit können die Orthofotos georeferenziert werden.

Abb. 2 zeigt einen Ausschnitt einer UAS-Aufnahme über Baltrum aus dem Jahr 2021. Die auf dem Boden sitzenden Vögel sind in der Regel gut erkennbar. Auffallend ist, dass die Vögel auf hellem, sandigen Boden (im linken Teil der Abbildung) dichter angesiedelt sind als auf den Vegetationsflächen (im rechten Teil der Abbildung). Es handelt sich dabei um die Vogelarten Brandseeschwalbe und Lachmöwe, die verschiedene Bruthabitate präferieren.

Tab. 1: Details zu den Untersuchungsgebieten und genutzten Aufnahmesystemen

	Spiekeroog	Baltrum
Gebietsgröße	1,74 km <sup>2</sup>	0,34 km <sup>2</sup>
Aufnahmezeitpunkt	29.05.2018	28.05.2021
UAS	Delair UX11	DJI Matrice 300 RTK
Flughöhe	75 m	65 m
Bodenauflösung	10,0 mm pro Pixel	7,5 mm pro Pixel



Abb. 2: Orthofotoausschnitt über Baltrum mit Brutkolonien der Brandseeschwalbe und der Lachmöwe

## 2.2 Manuelle Datenauswertung

Die Orthofotos werden mittels manueller, visueller Analyse von Experten des NLWKNs ausgewertet. Dabei wird jeder Seevogel im Orthofoto markiert und seiner entsprechenden Art zugeordnet. Die Vogelarten, die in den erfassten Brutgebieten nisten sind: Der Austernfischer, die Brandseeschwalbe, die Gans, die Heringsmöwe, die Lachmöwe, der Löffler, die rotfüßige Seeschwalbe und die Silbermöwe. Letztendlich liegt es im Ermessen der auswertenden Person, welcher Art ein Individuum zugeordnet wird. Damit ist die zugewiesene Vogelart mit einer Unsicherheit behaftet. Das Vorkommen der unterschiedlichen Vogelarten ist Standort spezifisch. Während im Brutgebiet auf Spiekeroog vor allem Heringsmöwen und Silbermöwen zu finden sind, brüten auf Baltrum vorwiegend Brandseeschwalben und Lachmöwen.

Der manuelle Detektions- und Klassifikationsprozess für das Brutgebiet auf Spiekeroog nimmt ca. 1,5 Wochen in Anspruch. Für das kleinere Brutgebiet auf Baltrum, werden ca. 3 Tage benötigt. Das Ergebnis der manuellen Auswertung dient in dieser Arbeit als Grundlage für die Generierung eines Datensatzes für das Training und die Evaluierung lernbasierter Klassifikatoren, welche Bildausschnitte als *Hintergrund* oder als entsprechenden *Seevogel* einordnen. Dazu werden aus den Orthofotos Bildausschnitte fester Größe extrahiert, die jeweils einen Seevogel abbilden. Der Seevogel liegt dabei im Zentrum des Bildausschnittes und hat eine zufällige Ausrichtung. Zusätzlich werden Bildausschnitte erstellt, welche ausschließlich Hintergrund, sprich keinen Seevogel, abbilden. Um einen möglichst fehlerfreien Trainingsdatensatz zu gewährleisten, werden alle Bildausschnitte und ihre entsprechende Klassenzuordnung händisch kontrolliert. Dabei werden fehlerhaft gekennzeichnete Bildausschnitte oder Bildausschnitte, die Artefakte aufweisen, aussortiert. Die Klassenverteilungen der resultierenden Datensätze sind in Tab. 2 aufgeschlüsselt.

Tab. 2: Art und Anzahl verschiedener Seevögel in den erstellten Datensätzen

Vogelart	Abkürzung	Vorkommen auf Spiekeroog	Vorkommen auf Baltrum
Austernfischer	AF	81	80
Brandseeschwalbe	BS	-	3.261
Gans	G	900	97
Heringsmöwe	HM	9.450	704
Lachmöwe	LM	149	24.300
Löffler	LF	132	60
Rotfüßige Seeschwalbe	RS	-	104
Silbermöwe	SM	2.110	524
Hintergrund	HG	4.336	6.122

### 3 Methodik

Für die Unterteilung von Bildausschnitten in die Klassen *Seevogel* und *Hintergrund* werden zwei lernbasierte Methoden - SVM und CNN - angewendet und verglichen. Während die SVM lernt, Bildausschnitte anhand extrahierter Histogram of Oriented Gradients (HOG)-Deskriptoren zu klassifizieren, lernt das CNN sowohl die Extraktion geeigneter Merkmale als auch die darauf basierende Klassifikation. Im Folgenden werden beide Ansätze näher erläutert.

#### 3.1 Klassifikation mittels SVM

Um charakteristische Bildinformation einzelner Bildausschnitte in kompakter Form darzustellen, kommen in der Regel Merkmalsdeskriptoren zum Einsatz. Zu den gewünschten Eigenschaften eines Deskriptors (auch Merkmal genannt) gehört die Invarianz gegenüber Skalierung, Translation und Rotation, sowie die Robustheit in Hinblick auf Beleuchtungsveränderung und Bildrauschen. In dieser Arbeit wird der sogenannte HOG-Deskriptor verwendet, welcher auf Helligkeitsgradienten innerhalb von Pixelnachbarschaften basiert (DALAL & TRIGGS 2005). Für die Berechnung des Deskriptors werden die RGB-Bilder zunächst in Grauwertbilder umgewandelt und mittels Sobel-Operator werden für jedes Pixel Gradientenbetrag und -orientierung bestimmt. Darauffolgend wird das Bild in mehrere, rechteckige Bereiche gleicher Größe unterteilt, die als Zellen bezeichnet werden. In dieser Arbeit wird dazu eine Zellengröße von  $9 \times 9$  Pixel verwendet. Für jede dieser Zellen wird ein Histogramm berechnet, welches den akkumulierten Gradientenbetrag über die Gradientenorientierung aufträgt. Dies ist in Abb. 3 veranschaulicht. Um den Einfluss von Beleuchtungsänderung innerhalb eines Bildes zu reduzieren, wird eine Normalisierung der Histogrammwerte über einen größeren Bildbereich (Block) durchgeführt. In dieser Arbeit wird dazu eine Blockgröße von  $2 \times 2$  Zellen genutzt. Der HOG-Deskriptor für den gesamten Bildausschnitt ergibt sich aus der Verkettung aller normalisierten Blockhistogramme.



Die HOG-Deskriptoren bilden die Eingangsdaten für die SVM. Innerhalb des Trainingsprozesses wird anhand von Beispielen eine Hyperebene gesucht, welche die HOG-Deskriptoren unterschiedlicher Klassen bestmöglich trennt. Da nicht davon ausgegangen wird, dass die vorliegenden HOG-Deskriptoren unterschiedlicher Klassen linear trennbar sind, wird der sogenannte Kernel-Trick angewendet, wodurch die Merkmale in einen höherdimensionalen Raum transformiert werden, in welchem eine passende Trennebene einfacher gefunden wird. Nach dem Training, wird die SVM genutzt, um ungesehene Bildausschnitte anhand ihrer HOG-Deskriptoren einer Klasse zuzuweisen.

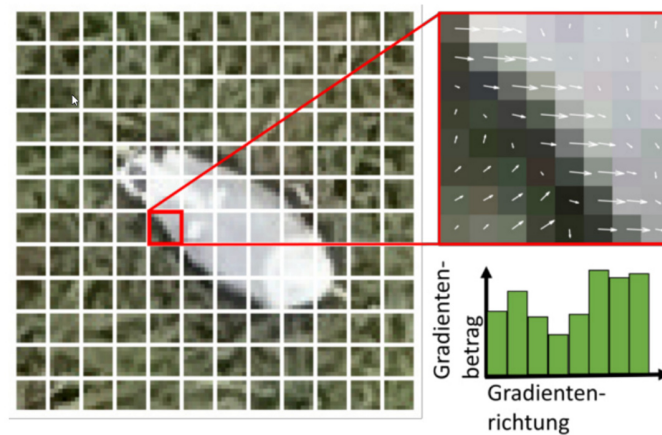


Abb. 3: Erstellung von HOG-Deskriptoren durch Histogrammbildung basierend auf Gradientenbetrag und -orientierung

### 3.2 Klassifikation mittels CNN

Im Gegensatz zu der SVM-basierten Klassifikation entfällt bei der CNN-basierten Klassifikation der Schritt der Merkmalsextraktion, da die RGB-Bilder direkt als Eingangsdaten genutzt werden können. Der Aufbau des CNNs, welches in dieser Arbeit genutzt wird, ist in Abb. 4 dargestellt. Der Bildausschnitt, der als Eingabeschicht dient, wird in der Convolutional-Schicht zunächst mit unterschiedlichen Filterkernen gefaltet. Die Filtergewichte stellen dabei variable Parameter dar, die im Rahmen des Trainingsprozesses optimiert werden. Die resultierenden Merkmale werden normiert (Batch Normalisierung) und es folgt eine nichtlineare Aktivierungsfunktion (ReLU). Durch eine Max-Pooling-Schicht werden die Merkmalsinformationen räumlich aggregiert und anschließend in den vollvernetzten Schichten miteinander verknüpft. Die Softmax-Funktion am Ende des Netzes transformiert die Ausgabe der vorangegangenen Schicht in eine kategoriale Wahrscheinlichkeitsverteilung. Die Optimierung der Netzparameter erfolgt anhand von Trainingsdaten durch Minimierung der Kreuzentropiefunktion. Nach der Trainingsphase können ungesehene Bildausschnitte vorwärts durch das Netz propagiert werden und einer der zuvor definierten Klassen zugewiesen werden.

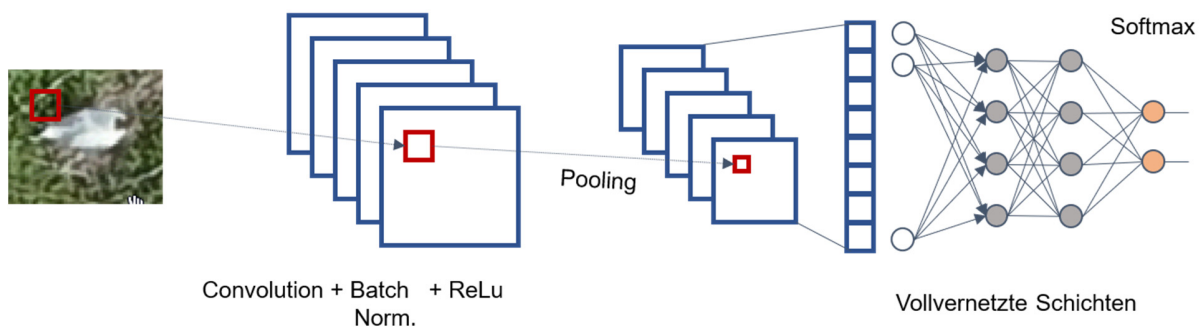


Abb. 4: Architektur des eingesetzten CNNs

### 3.3 Unausgewogene Trainingsdaten

Wie der in Tab. 2 beschriebenen Klassenverteilung des vorliegenden Datensatzes zu entnehmen ist, unterscheidet sich die Anzahl von Trainingsbeispielen pro Vogelart stark. Lernbasierte Modelle, die auf unausgewogenen Daten trainiert werden, neigen dazu überrepräsentierte Klassen aufgrund ihrer höheren Eintrittswahrscheinlichkeit häufiger vorherzusagen, während unterrepräsentierte Klassen teilweise vollständig ignoriert werden. Um dem ungewünschten Effekt einer Klassenüberanpassung entgegen zu wirken, werden im Rahmen der CNN-Klassifikation zwei Gegenmaßnahmen eingesetzt und verglichen. Zum einen werden die Trainingsdaten mittels einer Kombination aus Datenerweiterung und -reduktion ausbalanciert, zum anderen wird eine robuste Fehlerfunktion, der sogenannte Focal Loss, eingesetzt.

#### 3.3.1 Datenerweiterung und -reduktion

Unter Datenerweiterung versteht man die synthetische Vergrößerung von Datensätzen. Dieses Verfahren wird häufig eingesetzt, wenn eine unzureichende Menge an Daten für das Training eines Klassifikators zur Verfügung steht. Aus den vorhandenen Daten werden dabei beispielsweise durch Translation, Rotation und Skalierung synthetische Daten generiert. In dieser Arbeit wird diese Datenerweiterung genutzt, um synthetische Daten für die unterrepräsentierten Klassen zu erstellen. Damit ein Klassengleichgewicht erreicht werden kann ohne einen unverhältnismäßig hohen Anteil synthetischer Daten zu verwenden, wird die Menge der Daten überrepräsentierter Klassen durch eine zufällige Auswahl reduziert.

#### 3.3.2 Focal Loss

Wird für das Training des zuvor vorgestellten CNNs die Kreuzentropie-Fehlerfunktion verwendet, wird der Gesamtfehler durch überrepräsentierte Klassen in einem unausgewogenen Datensatz dominiert. Dadurch werden die trainierbaren Parameter in die Richtung aktualisiert, dass das Modell mehr Vertrauen in die Vorhersage der überrepräsentierten Klassen gewinnt, während die Minderheitsklassen weniger Gewicht bekommen. Die Focal Loss-Funktion, die von LIN et al. (2017) vorgeschlagen wird, stellt eine Modifikation der Kreuzentropie-Fehlerfunktion dar, die dazu führt, dass der Fokus auf schwierige, fehlklassifizierte Beispiele gesetzt wird. Dazu wird ein dynamischer Skalierungsfaktor eingeführt, welcher mit zunehmendem Vertrauen des Modells in die richtige Klasse gegen Null sinkt.

## 4 Ergebnisse

Für das Training und die Evaluation der Klassifikatoren werden die beiden Datensätze (Spiekeroog und Baltrum) in einem Verhältnis von 2:3 in Trainings- und Testdaten aufgeteilt. Das Training und die Evaluation erfolgen für beide Datensätze getrennt. Der Vergleich von SVM- und CNN-basierter Klassifikation wird anhand einer binären Klassifikation ausgewertet, bei welcher ein Bildausschnitt der Klasse *Seevogel* oder *Hintergrund* zugewiesen wird. Die Auswertung der Multiklassen-Klassifikation, bei welcher zusätzlich die Vogelart bestimmt wird, beschränkt sich auf die CNN-basierte Klassifikation. Für die Darstellung und Bewertung der Ergebnisse, die auf den Testdaten erzielt werden, wird jeweils die Konfusionsmatrix herangezogen sowie die daraus abgeleiteten Kenngrößen recall, precision, Gesamtgenauigkeit und balanced accuracy.

### 4.1 Binäre Klassifikation

Die Ergebnisse der binären Klassifikation mittels SVM und CNN sind in Tab. 3 (Spiekeroog) und Tab. 4 (Baltrum) dargestellt. In der letzten Zeile sind jeweils die recall-Werte, in der letzten Spalte die precision-Werte und in der rechten unteren Ecke die Gesamtgenauigkeit aufgetragen. Beide Methoden erzielen mit über 95% sehr hohe recall-Werte für die Klasse *Seevogel* sowie sehr hohe precision-Werte (>85%), wobei der CNN-Klassifikator um wenige Prozentpunkte besser abschneidet. Größere Unterschiede zwischen den beiden Methoden zeigen sich in der Erkennung der Klasse *Hintergrund*. Während die SVM 81,1% bzw. 88,9% der *Hintergrund*-Bilder als solche erkennt, erzielt das CNN recall-Werte von 91,4% bzw. 96,1%. Die Ergebnisse zeigen, dass die CNN-basierte Methode der SVM überlegen ist. Die durch das CNN automatisch gelernten Merkmale scheinen für die Unterscheidung von *Seevogel* und *Hintergrund* somit besser geeignet zu sein als die HOG-Merkmale.

Tab. 3: Testergebnisse, die auf dem Spiekeroog-Datensatz mittels SVM- und CNN-basierter Klassifikation erzielt werden

		Referenz		
		<i>Hintergrund</i>	<i>Seevogel</i>	
Prädiktion	<i>Hintergrund</i>	2.138	366	0,85
	<i>Seevogel</i>	498	7.982	0,94
		0,81	0,95	0,92

SVM-basierte Klassifikation

		Referenz		
		<i>Hintergrund</i>	<i>Seevogel</i>	
Prädiktion	<i>Hintergrund</i>	2.410	158	0,94
	<i>Seevogel</i>	226	8.190	0,97
		0,91	0,98	0,96

CNN-basierte Klassifikation



Tab. 4: Testergebnisse, die auf dem Baltrum-Datensatz mittels SVM- und CNN-basierter Klassifikation erzielt werden

		Referenz							
		<i>Hintergrund</i>	<i>Seevogel</i>						
Prädiktion	<i>Hintergrund</i>	3.266	415	0,89	Prädiktion	<i>Hintergrund</i>	3.529	94	0,97
	<i>Seevogel</i>	407	16.640	0,98		<i>Seevogel</i>	144	16.961	0,99
		0,89	0,98	0,96			0,96	0,98	0,99
SVM-basierte Klassifikation					CNN-basierte Klassifikation				

## 4.2 Multiklassen-Klassifikation

Bei der Multiklassen Klassifikation wird zusätzlich zwischen verschiedenen Vogelarten unterschieden. Für den Spiekeroog-Datensatz erfolgt eine Einteilung in die sieben Klassen *Hintergrund* (HG), *Heringsmöwe* (HM), *Silbermöwe* (SM), *Austernfischer* (AF), *Gans* (G), *Löffler* (LF) und *Lachmöwe* (LM). Die im Baltrum-Datensatz abgebildeten Seevögel werden in die Klassen *Brandseeschwalbe* (BS), *Lachmöwe* (LM) und *andere Arten* (AA) unterteilt. Die Zusammenfassung unterschiedlicher Arten in die Klasse AA erfolgt aufgrund der niedrigen Anzahl verfügbarer Beispielfelder pro Klasse.

Für die CNN-basierte Multiklassen-Klassifikation kann ein ein- oder mehrstufiges Verfahren genutzt werden. Im einstufigen Verfahren, wird ein Modell trainiert, welches direkt zwischen sieben bzw. vier Klassen unterscheidet, welches die verschiedenen Vogelarten sowie die Klasse Hintergrundklasse HG beinhalten. Bei der mehrstufigen Klassifikation werden hingegen zwei Modelle trainiert. Das erste Modell wird für die binäre Klassifikation (*Seevogel*, *Hintergrund*) eingesetzt. Alle Bildausschnitte, die durch das erste Modell als *Seevogel* klassifiziert werden, werden von einem zweiten Modell der abgebildeten Vogelart zugeordnet. Durchgeführte Versuchsreihen haben gezeigt, dass sich die Klassifikationsleistung des einstufigen Verfahrens im Vergleich zu dem zweistufigen Verfahren nur geringfügig verschlechtert oder in manchen Fällen sogar verbessert. Da der Trainings- und Testaufwand des einstufigen Verfahrens deutlich geringer ist, wird dieses in den folgenden Analysen verwendet.

Die CNN-basierte Klassifikation der Vogelart wird in den folgenden drei verschiedenen Konfigurationen durchgeführt:

1. Training des CNNs mit Kreuzentropie-Fehlerfunktion anhand einer stratifizierten Stichprobe des jeweiligen Datensatzes.
2. Training des CNNs mit Kreuzentropie-Fehlerfunktion anhand ausbalancierter Trainingsdaten.
3. Training des CNNs mit Focal Loss-Funktion anhand einer stratifizierten Stichprobe des jeweiligen Datensatzes.

Tab. 5: Konfusionsmatrizen der CNN-basierten Klassifikation des Spiekeroog-Datensatzes

		Referenz								
		<i>HG</i>	<i>HM</i>	<i>SM</i>	<i>AF</i>	<i>G</i>	<i>LF</i>	<i>LM</i>		
Prädiktion	<i>HG</i>	2.493	102	16	6	86	4	2	0,92	Stratifizierte Trainingsdaten <b>Balanced accuracy: 0,60</b>
	<i>HM</i>	34	5.183	294	10	228	0	65	0,89	
	<i>SM</i>	10	272	931	0	16	9	17	0,74	
	<i>AF</i>	5	6	0	20	53	0	0	0,24	
	<i>G</i>	21	62	1	13	155	0	1	0,61	
	<i>LF</i>	39	0	4	0	0	66	0	0,61	
	<i>LM</i>	0	45	19	0	2	0	5	0,07	
		0,96	0,91	0,74	0,42	0,29	0,83	0,06	0,83	
Prädiktion	<i>HG</i>	2.508	261	38	5	112	8	2	0,85	Ausbalancierte Trainingsdaten <b>Balanced accuracy: 0,62</b>
	<i>HM</i>	26	4.105	197	2	140	0	40	0,91	
	<i>SM</i>	10	703	945	0	24	6	21	0,55	
	<i>AF</i>	0	23	0	23	68	0	0	0,20	
	<i>G</i>	29	215	4	19	192	1	1	0,42	
	<i>LF</i>	26	6	6	0	0	64	0	0,63	
	<i>LM</i>	3	357	76	0	4	0	25	0,05	
		0,96	0,72	0,75	0,46	0,36	0,81	0,28	0,75	
Prädiktion	<i>HG</i>	2.426	51	16	2	47	9	1	0,95	Stratifizierte Trainingsdaten & Focal Loss-Funktion <b>Balanced accuracy: 0,60</b>
	<i>HM</i>	76	5.202	373	9	242	1	65	0,87	
	<i>SM</i>	23	286	829	0	21	5	12	0,70	
	<i>AF</i>	5	0	0	23	49	0	0	0,30	
	<i>G</i>	23	74	1	15	179	0	0	0,61	
	<i>LF</i>	44	0	4	0	0	64	0	0,57	
	<i>LM</i>	5	57	43	0	2	0	11	0,09	
		0,93	0,92	0,65	0,46	0,33	0,81	0,13	0,82	

Tab. 6: Konfusionsmatrizen der CNN-basierten Klassifikation des Baltrum-Datensatzes

		Referenz					Referenz					
		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>		
Prädiktion	<i>HG</i>	3515	16	58	72	0,96	<i>HG</i>	3541	16	102	138	0,93
	<i>BS</i>	59	1691	235	62	0,83	<i>BS</i>	77	1767	656	87	0,68
	<i>LM</i>	18	225	14068	317	0,96	<i>LM</i>	7	151	13210	154	0,98
	<i>AA</i>	81	25	219	489	0,60	<i>AA</i>	48	23	612	561	0,45
		0,96	0,86	0,96	0,52	0,93	0,96	0,90	0,91	0,60	0,91	

Stratifizierte Trainingsdaten                      Ausbalancierte Trainingsdaten

**Balanced accuracy: 0,83**                      **Balanced accuracy: 0,84**

		Referenz				
		<i>HG</i>	<i>BS</i>	<i>LM</i>	<i>AA</i>	
Prädiktion	<i>HG</i>	3442	57	87	82	0,94
	<i>BS</i>	121	1680	365	80	0,75
	<i>LM</i>	55	204	13953	400	0,95
	<i>AA</i>	55	16	175	378	0,61
		0,94	0,86	0,96	0,40	0,92

Stratifizierte Trainingsdaten & Focal Loss-Funktion

**Balanced accuracy: 0,78**

Die auf den Testdaten erzielten Ergebnisse der drei Ansätze sind in Form von Konfusionsmatrizen in Tab. 5 für den Spiekeroog-Datensatz und in Tab. 6 für den Baltrum-Datensatz dargestellt. Zusätzlich ist jeweils der balanced accuracy-Wert angegeben, welcher sich aus der Mittelung der klassenspezifischen recall-Werte ergibt und für die Bewertung der Ergebnisse nicht ausbalancierter Testdaten geeignet ist. Für den Spiekeroog-Datensatz zeigt sich, dass der Hintergrund (*HG*) sowie die Klassen *HM* und *SM* mit recall-Werten von 96%, 91% und 74% sowie mit hohen precision-Werten (>75%) zuverlässig klassifiziert werden. Wie Tab. 2 zu entnehmen ist, sind diese Klassen in den zugrunde liegenden Trainingsdaten gut repräsentiert. Der Löffler (*LF*), der durch seine markante helle Federung eine homogene und gut abgegrenzte Klasse darstellt, wird trotz weniger Trainingsbeispiele gut erkannt. Die Klassen *AF*, *G* und insbesondere *LM*, die in den Trainingsdaten unterrepräsentiert sind, werden hingegen nur selten richtig erkannt und häufig der überrepräsentierten Klasse *HM* zugeordnet. Insgesamt ergibt sich eine Gesamtgenauigkeit von 83% und ein balanced accuracy-Wert von 60%. Wird das Training mit ausbalancierten Daten durchge-

führt, werden die zuvor unterrepräsentierten Klassen (*AF*, *G* und *LM*) insgesamt häufiger prädiert. Die recall-Werte verbessern sich um 4%, 7% und 22%, liegen allerdings noch immer deutlich unter 50%. Gleichzeitig verschlechtert sich der recall-Wert der Klasse *HM*, für welche die Anzahl der Trainingsbeispiele reduziert wurde von 92% auf 72%. Insgesamt ergibt sich im Vergleich zu dem Einsatz stratifizierter Trainingsdaten ein leicht verbesserter balanced accuracy-Wert von 62%. Für die dritte Trainingskonfiguration (Focal Loss-Funktion und stratifizierte Stichprobe), ist die Verbesserung der recall-Werte unterrepräsentierter Klassen schwächer ausgeprägt. Der balanced accuracy-Wert wird nicht verbessert und liegt bei 60%.

Die Auswertung der Klassifikationsergebnisse für den Baltrum-Datensatz (Tab. 6) bestätigt, dass die Nutzung ausbalancierter Daten zu dem höchsten balanced accuracy-Wert (84%) führt, während das Training mit der Focal Loss-Funktion (78%) gegenüber der Kreuzentropie-Fehlerfunktion (83%) insgesamt keinen Vorteil bringt. Für alle drei Trainingskonfigurationen zeigt sich, dass die beiden Hauptklassen *BS* und *LM* sehr gut voneinander getrennt werden können, wobei dies mit dem Training auf einer stratifizierten Stichprobe am besten gelingt. Problematisch ist jeweils die Klassifikation der Klasse *AA*. Als Grund hierfür lässt sich neben der vergleichsweise geringen Anzahl von Trainingsbeispielen, die hohe Heterogenität dieser Klasse, durch die Zusammenfassung unterschiedlicher Vogelarten, anführen.

## 5 Fazit und Ausblick

In dieser Arbeit wurden überwachte Lernverfahren eingesetzt und verglichen, welche das Potential bieten die Erkennung und Klassifikation von Seevögeln in UAS-Aufnahmen zu automatisieren. Für eine binäre Einteilung von Bildausschnitten in die Klassen *Seevogel* oder *Hintergrund* wurde eine SVM-basierte Klassifikation basierend auf HOG-Merkmalen und ein flaches CNN angewendet. Beide Ansätze liefern sehr gute Ergebnisse, wobei die Klassifikationsleistung des CNNs, mit Gesamtgenauigkeiten von bis zu 99%, der SVM, mit einer Gesamtgenauigkeit von bis zu 96%, leicht überlegen ist. Das CNN wurde des Weiteren für die Klassifikation von bis zu sechs verschiedenen Vogelarten eingesetzt. Dabei hat sich gezeigt, dass Vogelarten, für die eine gute Datengrundlage im Sinne umfangreicher Trainingsdaten verfügbar ist, zuverlässig erkannt und unterschieden werden können. In den Trainingsdaten unterrepräsentierte Vogelarten, die zu dem eine hohe Varianz innerhalb der Klasse aufweisen, werden von dem CNN-Modell seltener erkannt und in der Regel den stark repräsentierten Klassen zugewiesen. Um die recall-Werte unterrepräsentierter Klassen zu erhöhen, wurde das CNN zum einen mit einem ausbalancierten Datensatz und zum anderen unter Verwendung der Focal Loss-Funktion trainiert. Die Ergebnisse haben gezeigt, dass die Focal Loss-Funktion die bestehende Herausforderung nicht löst, während sich für das Training mit ausbalancierten Daten eine leichte Verbesserung der balanced accuracy-Werte ergibt. Anstelle der Datenerweiterung und -reduktion die in dieser Arbeit betrachtet wurde, empfiehlt sich für zukünftige Arbeiten unterrepräsentierte Klassen durch Hinzunahme weiterer UAS-Daten (z.B. aus Messkampagnen vergangener Jahre) zu stärken. Auf diese Weise ergeben sich mehr echte Trainingsdaten, welche die CNN-basierte Klassifikation robuster werden lassen. Ein weiterer Fokus zukünftiger Arbeiten besteht darüber hinaus in der Anwendung des Klassifikators auf das gesamte

Befliegungsgebiet. Durch den Abgleich von Klassifikationsergebnissen auf zwei zeitlichen versetzten Aufnahmen, lassen sich Brutvögel identifizieren und automatisch zählen.

## 6 Danksagung

Für die Beratung sowie die Bereitstellung und die manuelle Auswertung der in dieser Arbeit genutzten UAS-Aufnahmen möchten wir uns herzlich bei Holger Dirks und Birgit Hepner des NLWKN bedanken.

## 7 Literaturverzeichnis

- AFÁN, I., MÁÑEZ, M. & DIAZ-DELGADO, R., 2018: Drone Monitoring of Breeding Waterbird Populations: The Case of the Glossy Ibis. *Drones*, **2**(4), 42-55.
- DALAL, N. & TRIGGS, B., 2005: Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 886-893.
- DIAZ-DELGADO, R., MÁÑEZ, M., MARTÍNEZ, A., CANAL, D., FERRER, M. & ARAGONÉS, D., 2017: Using UAVs to Map Aquatic Bird Colonies. *The Roles of Remote Sensing in Nature Conservation*, 277-291.
- GRENZDÖRFFER, G., 2013: UAS-based Automatic Bird Count of a Common Gull Colony. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **40**(1/W2), 169-174.
- HODGSON, J., MOTT, R., BAYLIS, S., PHAM, T., WOTHERSPOON, S., KILPATRICK, A., RAJA SEGARAN, R., REID, I., TERAUDS, A. & KOH, L., 2018: Drones Count Wildlife More Accurately and Precisely than Humans. *Methods in Ecology and Evolution*, **9**(5), 1160-1167.
- HONG, S. J., HAN, Y., KIM, S. Y., LEE, A. Y. & KIM, G., 2019: Application of Deep-learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery. *Sensors*, **19**(7), 1651-1667.
- KELLENBERGER, B., VEEN, T., FOLMER, E., & TUIA, D., 2021: 21 000 Birds in 4.5 h: Efficient Large-scale Seabird Detection with Machine Learning. *Remote Sensing in Ecology and Conservation*, **7**(3), 445-460.
- LIN, T., GOYAL, P., GIRSHICK, R., HE, K. & DOLLAR, P., 2017: Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980-2988.
- RUSH, G. P., CLARKE, L. E., STONE, M. & WOOD, M. J., 2018: Can Drones Count Gulls? Minimal Disturbance and Semiautomated Image Processing with an Unmanned Aerial Vehicle for Colony-Nesting Seabirds. *Ecology and Evolution*, **8**(24), 12322-12334.
- SARDÀ-PALOMERA, F., BOTA, G., VINOLO, C., PALLARES O. SAZATORNIL, V., BROTONS, L., GOMARIZ, S. & SARDA, F., 2012: Fine-Scale Bird Monitoring from Light Unmanned Aircraft Systems. *Ibis*, **154**(1), 177-183.