

Wildtier-Monitoring mit UAVs – Künstliche Intelligenz zur automatisierten Detektion von Infrarot-Signaturen

ADRIAN F. MEYER¹, NATALIE LACK & DENIS JORDAN

Zusammenfassung: Die Detektion von Wildtieren ist ein zentrales Monitoring-Instrument in Ökologie, Jagd, Forst- und Landwirtschaft. Bisherige Verfahren sind aufwändig, basieren häufig lediglich auf indirekten Nachweisen und bieten so oft nur eine grobe Schätzung der Bestände. Die in dieser Arbeit durchgeführte fernerkundliche Auswertung von UAV-Befliegungen über dem Südschwarzwald und der Nordwestschweiz zeigte, dass sich insbesondere Wärmebilddaten für eine Automatisierung der Wildtier-Detektion eignen. Hierzu wurde ein modernes Verfahren der künstlichen Intelligenz (Faster R-CNN) aufgebaut, das in der Lage ist, durch Training Eigenschaftsmerkmale aus markierten Tiersignaturen zu extrahieren. Für einige Tierarten (Hirsche, Ziegenartige, Wisente, Weidevieh) konnten in der anschließenden Anwendung (Inferencing) extrem robuste Detektionsergebnisse erreicht werden. Die effiziente Implementierung des Prototyps erlaubt eine Echtzeitanalyse von Live-Video-Feeds unter Feldbedingungen. Mit einer Detektionsrate von 92,8% pro Tier, bzw. 88,6% in der Klassifikation nach Tierart, konnte so gezeigt werden, dass die neue Technologie ein enormes Innovationspotential für die Zukunft des Wildtier-Monitorings aufweist.

1 Einleitung

Für Anwendungsfelder wie Populationsmanagement, Rehkitzrettung und Wildschadensprävention in Ökologie, Jagd, Forst- und Landwirtschaft ist es von entscheidender Bedeutung, eine möglichst präzise Erfassung der Wildtierbestände vornehmen zu können. Beim konventionellen Monitoring kommen derzeit zumeist Methoden zum Einsatz, die jeweils entscheidende Nachteile mit sich bringen (SILVEIRA et al. 2003): Zählkampagnen mit visueller Bestätigung (Scheinwerfersuche auf Waldwegen) sind enorm personalintensiv; Kamerafallen-Analysen decken nur einen kleinen Ausschnitt der Landschaft ab; Jagdquoten und Wildunfallstatistiken sind mit einem starken Bias verbunden; Peilsender sind zwar sehr genau, aber auch invasiv und in der Umsetzung aufwändig. Das Institut Geomatik (FHNW) kooperiert seit Januar 2018 mit der *Stiftung Wildtiere* des *Aargauischen Jagdverbandes*, um ein Verfahren zur Wildtierdetektion mittels UAVs (*Unmanned Aerial Vehicles*) zu entwickeln. Es wird untersucht, in wie weit automatisierte Fernerkundungsmethoden Vorteile gegenüber dem konventionellen Monitoring bieten, indem zeitliche oder personelle Ressourcen eingespart und somit Erhebungen präziser und vollständiger durchgeführt werden können (GONZALEZ et al. 2016). Zentrale Fragen, welche dieses Studie beantworten soll, stellen sich dabei hinsichtlich der Wahl von Sensorik und Trägersystemen, der generellen Sichtbarkeit von Tiersignaturen auf Infrarotluftbildern (z.B. die Robustheit gegenüber Verschattungen im Mischwald), sowie der Struktur eines möglichst performanten Algorithmus zur automatisierten Detektion und Klassifikation der Wildtierindividuen. Ein Ergebnis dieser Analyse ist ein Prototyp, der eine automatisierte Tierdetektion auf Luftbilddaten ermöglichen soll.

¹ Fachhochschule Nordwestschweiz, Institut Geomatik, Hofackerstr. 30, CH – 4132 Muttenz, E-Mail: adrian.meyer@fhnw.ch

2 Methode

2.1 Datenerhebung

Im Frühjahr 2018 wurden 27 Befliegungen über sieben naturnahen Wildgehegen mit heimischen Tierarten in der Nordwestschweiz und dem Südschwarzwald durchgeführt. Für jedes Gehege wurden mit dem Multikopter *senseFly Albris* oder dem Fixed-Wing-UAV *senseFly eBee* ca. 500 RGB-Bilder, 500 NIR-Multispektralbilder und 5000 TIR-Wärmebilder (radiometrische Thermogramme) generiert, um einen Technologievergleich zu ermöglichen (siehe Abb. 1). Die Aufnahmezeit (Februar/März) wurde so gewählt, dass der Wärmekontrast zwischen Tierkörper und zu meist bewaldeter Umgebung möglichst hoch ausfallen würde. Gleichzeitig sollte durch die laubfreie Vegetation möglichst wenig Verschattung entstehen.

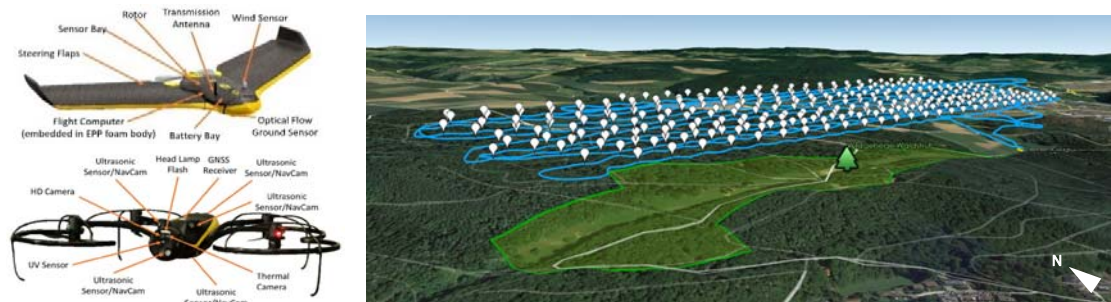


Abb. 1: Links: Eingesetzte senseFly-Fluggeräte „eBee“ (oben) und „Albris“ (unten). Rechts: Typische Flugbahn mit der eBee (blau) über einem Tierpark (grün) mit den Auslösepositionen für Luftbilder (weiss). (Visualisierungen: GILLINS et al. 2018; GOOGLE 2018; SENSEFLY 2018)

Mit dem Fixed-Wing können große Gebiete unkompliziert mit austauschbaren Sensoren (RGB, NIR, TIR) erfasst werden, u.a. mit einer hochauflösenden Thermalkamera (*ThermoMap*, 640×512Px; max. 22ha bei 15cm/Px GSD und 100m AGL). Mit dem Multikopter kann zwar aufgrund dessen Schwebefähigkeit wesentlich flexibler und tiefer geflogen werden, allerdings weist die fest verbaute Thermalkamera eine deutlich niedrigere Auflösung auf (80×60Px). Das laute Rotorengeräusch bei einer niedrigen Flugbahn stellt zudem im Vergleich zum Fixed-Wing einen wesentlich stärkeren Eingriff in das Tiervverhalten dar.

2.2 Vorprozessierung

Die sehr hoch aufgelösten RGB- und NIR-Aufnahmen (~3cm/Px GSD) eignen sich gut für das Erstellen von Übersichtskarten in Form von Orthophotomosaiken, weisen jedoch häufig keinen ausreichenden Kontrast für die visuelle Erkennung von Tiersignaturen unter laubfreier Vegetation auf. Dies konnte im weiteren Verlauf der Studie auch durch terrestrische hyperspektrale Referenzmessungen (λ : 350-1000nm) an Waldboden, Vegetation und Tierkörpern verifiziert werden.

Die Thermogramme hingegen zeigen kontrastreiche Signaturen einzelner Wildtiere (Abb. 2). Die Bilder eignen sich aber gleichzeitig kaum für den photogrammetrischen Bündelblockausgleich, da sich die Tiere in der Regel zwischen zwei Aufnahmen zu stark bewegen. In den relevanten Bildbereichen entsteht dadurch keine ausreichende Überlagerungstreue, sodass prozessierte TIR-Orthophotomosaiken zusammenhängender Habitate häufig keine sichtbaren Signaturen enthalten. Für

die automatisierte Analyse wurden die Thermogramme deshalb entweder direkt als nicht-orientierte Rohdaten weiterverarbeitet, oder einzeln per DSM-Projektion orthorektifiziert.

3 Analyse

3.1 Gestalt der thermalen Tiersignaturen

Sichtbare Änderungen im Erscheinungsbild der Signaturen wurden zunächst systematisch durch die Variation von Referenzparametern untersucht. So unterstützt eine flachere Aufnahmeperspektive die Tier-Identifikation durch einen menschlichen Betrachter (Abb. 2 links): Merkmale wie Kopf-Torso-Ratio oder Extremitäten treten deutlicher hervor. Die Abgrenzung der Individuen voneinander wird hingegen eher durch eine steilere Perspektive unterstützt.

In möglichst laubfreiem Mischwald reduziert dichtes Astwerk zwar teilweise den Signaturkontrast durch Konvektionswärmeverteilung und Abschirmung; Form, Umfang und die grundsätzliche Sichtbarkeit der Signaturen bleiben jedoch in weiten Teilen erhalten (Abb. 2 rechts).

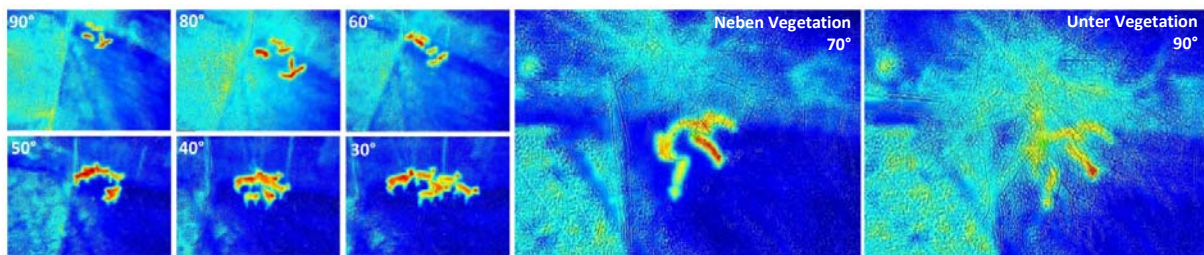


Abb. 2: Thermogramme mit den Signaturen eines Damhirsch-Rudels (sechs Tiere, Blau 4°C, Rot 10°C). Links: Signaturen aus sechs verschiedenen Aufnahme-Winkeln im Vergleich. Mitte/Rechts: Signaturen neben und unterhalb einer laubfreien Esche im Vergleich.

3.2 Strategien zur automatisierten Signaturdetektion

Mehrere Strategien zur automatisierten Detektion von Signaturen wurden iterativ implementiert und hinsichtlich ihrer Klassifikationsgenauigkeit und Anwendbarkeit überprüft. Der klassische fernerkundliche Ansatz, mittels objektbasierter Bildanalyse Thermogramme beispielsweise in *Erdas Imagine Objective* zu klassifizieren, wurde dabei verworfen. Mit dieser Methode konnte aufgrund der Vielgestaltigkeit der Signaturen kein merkmalsbeschreibender Satz von Variablen gefunden werden, der eine Detektionspräzision von über 41% erreichen würde. Gefaltete neuronale Netzwerke (Convolutional Neural Networks, CNN) hingegen zeigten in den letzten Jahren eine außerordentliche Robustheit bei der Bildklassifikation mittels automatischer Merkmalsextraktion (SZEGEDY et al. 2016). In den Abschnitten 3.3 und 3.4 werden zwei CNN-Ansätze beschrieben, die auf unterschiedliche Art und Weise eine präzise Tierdetektion erreichen.

3.3 Rastersegment-Klassifikation mit dichotomen CNN

Ein dichotomes („Zwei-Wege-Entscheidung“) CNN mit einer Tiefe von 7 Neuronen-Layern (Abb. 3 Mitte) wurde mit *Keras* und *Tensorflow* unter *Python 3.6* aufgebaut. Es klassifiziert Rastersegmente orthorektifizierter Thermogramme per Inferencing in die Klassen „Tier“ und „Nicht-Tier“. Als Input-Layer wird eine 64×64Px-Matrix verwendet, welche bei maximal möglicher GSD

geoprozessierten $5 \times 5\text{m}$ -Segmenten entspricht (Abb. 3 links). Nach ca. 3h Training auf Desktophardware kann für eine spezifische Befliegung eine hohe Klassifikationspräzision von ca. 90% erreicht werden (Abb. 3 rechts). Das Vorprozessieren der Thermaldaten (3D-Projektion auf DSM, Orthophoto-Generierung, Geoprozessierung) ist jedoch sehr zeit- und rechenintensiv und somit unter Feldbedingungen als nicht praktikabel automatisierbar einzustufen. Gerade bei zeitkritischen Anwendungen wie der Rehkitzrettung müssen Klassifikationsergebnisse im besten Fall bereits während des Flugs vorliegen. Das Inferencing auf Live-Rohdaten wäre diesen Limitationen nicht unterworfen. Durch die Rohdaten-Auflösung von $640 \times 512\text{Px}$ bietet dieser Ansatz dem UAV-Operator bei der praxisnahen Anwendung aufgrund der $64 \times 64\text{Px}$ -Inputauflösung jedoch nur ein grobes 10×8 -Detektionsraster.

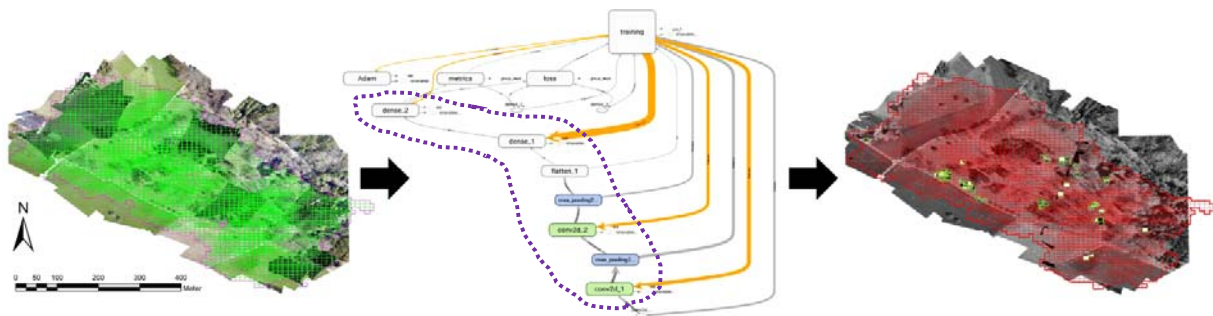


Abb. 3: Links: Ca. 10'000 $5 \times 5\text{m}$ Footprints als Inputkacheln, generiert aus 45 orthorektifizierten Thermo-grammen. Mitte: Schema des dichotomen neuronalen Netzwerks, Neuronenlayer in der violetten Markierung. Rechts: Klassifikation - 71 Kacheln Grün: „Tier“; Rest Rot: „Nicht-Tier“.

3.4 Objekterkennung mittels R-CNN

Für die Live-Rohdaten-Interpretation sind *Faster Region-based Convolutional Neural Networks* (*Faster R-CNN*) besser geeignet. Modelle dieser Klasse können Objekte auf höher aufgelösten Gesamtbildern klassifizieren, indem *Regions of Interest (RoI)* durch iteratives *Region Proposal* lokalisiert werden. Auch können gleichzeitig verschiedene Klassen trainiert und erkannt werden. Eingesetzt wird ein *Inception-v2*-Netzwerk (Vgl. Abb. 4), das mit einer Tiefe von 42 Neuronen-Layern den Aufbau der Pyramidenzellen im visuellen Cortex von Wirbeltieren nachahmt. Durch das Vortraining mit 100'000 Alltagsbildern (sog. *COCO-Dataset*) können die Kantengewichte zwischen den Neuronen-Layern im spezifischen Training schneller und effizienter an neue Ziele für das Festlegen der *Bounding Boxes* angepasst werden. Das Modell gilt auch bei teilweise limitierten Hardware-Anforderungen noch als schnell und präzise (SZEGEDY et al. 2016).

Die Implementierung erfolgte mittels der *Tensorflow Object Detection Library* mit Unterstützung durch das *Nvidia CUDA/cuDNN* Deep-Learning-Framework zur Parallelisierung der GPU-Shaderkerne. Für das Training wurde ein Testdatensatz von ca. 600 Thermalbildern mit ca. 8'000 Tiersignaturen durch Zeichnen von ca. 1'800 Bounding Boxes manuell markiert. Nach ca. 12 Stunden Training (ca. 100'000 Steps) wurde der ca. 50 Mbyte große *Frozen Inference Graph* exportiert. Ein performanter, auf Python-Scripts basierender Prototyp wendet dieses Wissensschema per Inferencing auf neue Thermaldaten an.

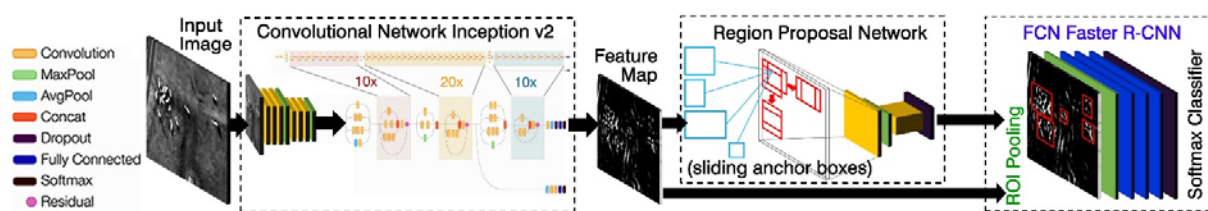


Abb. 4: Schematische Struktur des aufgebauten R-CNN (Subschema „Inception v2“ aus ALEM1, 2016)

4 Ergebnis

Im Vergleich erwies sich die Objekterkennung mittels R-CNN aufgrund der Fähigkeiten, Rohdaten verwenden und mehrere Klassen gleichzeitig trainieren zu können, als der überlegene Ansatz. In der prototypischen Umsetzung wurde deshalb diese Architektur verwendet.

Wird das Netzwerk nur auf die allgemeine Erkennung von Tieren trainiert (Abb. 5-A), so wird im Inferencing eine enorm hohe Detektionsquote von 92,8% pro Tier erreicht. Die Analysen zeigen sich bei der allgemeinen Detektion zudem relativ robust gegenüber Qualitätseinbußen bei den Eingabedaten (z.B. Bewegungsunschärfe und Verschattungen), da die Tiersignaturen zumeist in mindestens einem Frame korrekt erkannt werden und so für die Erfassung genutzt werden können.

	Allgemeine Detektion	Klasse Damwild*	Klasse Rotwild	Klasse Ziegen- artige
Detektierbarkeit pro Tier (mindestens in einem Frame)	92,8 %	96,9 %	84,2 %	93,8 %
Detektierbarkeit pro Einzelframe	82,7 %	76,2 %	70,8 %	65,9 %
Anteil Falsch- Positive an allen Detektionen in der gleichen Klasse	0,5 %	21,7 %	6,4 %	0,9 %
Anzahl Signaturen in der Auswertung	2 576	1 347	1 032	1 457
Anzahl Bounding Boxes im Training	1 628	647	353	323

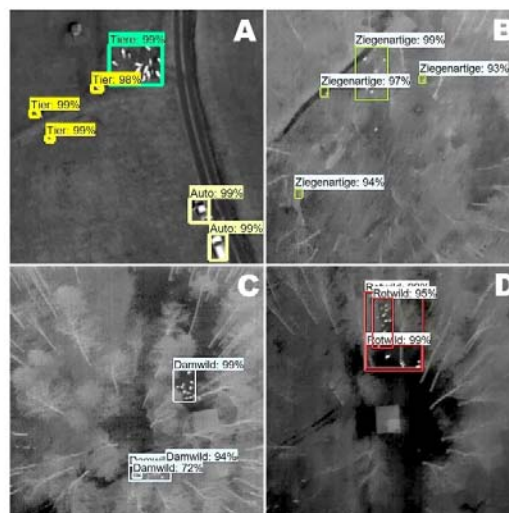


Abb. 5: Auswertung der Detektionsergebnisse auf Video-Feeds, welche einen Live-Überflug simulieren. Links: Tabelle der Zählstatistiken (*Klasse Damwild enthält Damwild, Sikawild und Axiswild). Der Anteil Falsch-Negativer Detektionen ist nicht aufgeführt, da er dem Kehrwert der Detektierbarkeit entspricht. Rechts: Beispiele per Inferenz errechneter Bounding Boxes.

Bei der Klassifikation nach Tierart fällt die Präzision leicht ab (Abb. 5-BCD), übertrifft jedoch für die Klassen *Damwild*, *Rotwild* und *Ziegenartige* die Erfolgsraten konventioneller Erfassungsmethoden noch immer bei weitem. *Wisente* und *schottische Hochlandrinder* erreichen ebenfalls eine sehr hohe Präzision (Detektierbarkeit >90% pro Tier, >80% pro Frame bei ≈ 150 trainierten Bounding Boxes). Diese Werte sind jedoch nicht mit denjenigen von Wildtieren in naturnahen Mischwald-Gehegen vergleichbar, da sowohl für das Training, als auch für das Inferencing nur Daten in offener Weidelandschaft zur Verfügung standen (Vgl. Weide in Abb. 5-A; Mischwald in

Abb. 5-BCD). Die Klassen *Schwarzwild*, *Menschen* und *kleinere Säugetiere* erreichten aufgrund einer zu niedrigen Anzahl an Trainingsdaten ($n < 60$) keine ausreichende Klassifikationspräzision.

5 Diskussion und Ausblick

UAV-basierte Infrarot-Thermographie mit State-of-the-Art-Verfahren des Deep Learnings zu verbinden, weist auf das Potential hin, Effizienz- und Qualitätssteigerungen bei der Populationsschätzung zu erreichen. Der aktuelle Standard – ein aufwändiger, personalintensiver Prozess der Scheinwerfer-Suche, bei welchem viele Kilometer Waldwege abgefahren werden, um nur einen kleinen, unbekannten Anteil der Tiere zu kartieren – ließe sich so mit modernen Methoden der Mustererkennung ergänzen. Der implementierte Prototyp erreicht auf mobiler Hardware (2016 Consumer-Grade Laptop) eine Inferencing-Performanz von ca. 8 FPS. Damit ist das System so effizient, dass es während des Flugs auf einen thermalen Live-Videofeed angewandt werden kann. Diese vielversprechenden Ergebnisse zeigen somit, dass eine Ablösung der klassischen Erfassungsmethoden für bestimmte Gebiete in Zukunft denkbar ist.

Ein weiterer wichtiger Anwendungsfall ist die Rehkitzrettung. In Mähwiesen versteckte Rehkitze fallen aufgrund ihres Drückreflexes häufig den Mähdreschern zum Opfer. Werden heute Thermal-UAVs eingesetzt, läuft der Prozess jedoch noch weitgehend manuell ab. Zudem ist das Training der Piloten zur Signatur-Erkennung aufwändig. Die vorgestellte Softwareautomatisierung kann die UAV-basierte Rehkitzrettung in Zukunft daher erheblich verfügbarer machen.

Bei der Wildschadensprävention stehen zumeist Wildschweinrotten im Fokus, die sich von außen unsichtbar einen Rückzugsort in Ackerkulturen trampeln. Die Verortung der Tiere ist mit dieser Technologie noch vor der Entstehung eines größeren Schadens möglich. Sowohl für die Wildschadensprävention, als auch für die Rehkitzrettung bedarf es zum operativen Einsatz noch zusätzlicher Trainingsdaten. Sind diese erhoben und markiert, kann das bestehende Deep-Learning-Netzwerk per *Fine Tuning* weitertrainiert werden und so auf dem bereits erworbenen Wissensschatz aufgebaut werden.

Durch die schnellen Fortschritte in der UAV-Technologie ist es durchaus denkbar, dass kleinere Multikopter bald leiser fliegen können und somit weniger Störungen des Tierverhaltens auslösen. Mit einer niedrigeren Flughöhe und stärkeren Sensoren wären diese in der Lage, noch bessere Thermogramme zu generieren, was wiederum die Signaturklassifikation für das neuronale Netzwerk erleichtert. Denkbar wäre die Identifikation weiterer Individualmerkmale der bereits analysierten Arten wie Alter und Geschlecht, oder aber auch eine Ausweitung der Analyse auf kleinere Spezies wie Dachse, Feldhasen und Füchse, sowie seltene Spezies wie Luchse und Wölfe.

6 Literaturverzeichnis

- ALEMI, A., 2016: Improving Inception and Image Classification in Tensorflow. Google AI Blog.
 GOOGLE, 2018: Google Earth Pro 7.3.1, Luftbild-Textur: GeoBasis-DF/BKG, 2017-08-07.
 GONZALEZ, L., MONTES, G., PUIG, E., JOHNSON, S., Mengersen, K. & Gaston, K., 2016: Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation. *Sensors*, **16** (1), Beitrag 97.

- GILLINS, D., PARRISH, C., GILLINS, M., H. & SIMPSON, C., 2018: Eyes in the Sky: Bridge Inspections with Unmanned Aerial Vehicles. Oregon Dept. of Transportation, SPR 787 Final Report.
- SENSEFLY, 2018: <https://www.sensefly.com/drone/ebec-mapping-drone/> (6.5.2018).
- SILVEIRA, L., JACOMO, A. & DINIZ-FILHO, J., 2003: Camera trap, line transect census and track surveys: a comparative evaluation. *Biological Conservation*, **114** (3), 351-335.
- SZEGEDY, C., VANHOUCKE, V., IOFFE, S., SHLENS, J. & WOJNA, Z., 2016: Rethinking the inception architecture for computer vision. *IEEE Conference on Computer Vision and Pattern Recognition*, 2818-2826.