DeepBirdDetect

Automatic Bird Detection of Endangered Species Using Deep Neural Network

Christoph Scholz, Katharina Brauns, Oliver Haddenhorst, Mira Jürgens

Kontakt: Christoph Scholz | +49 160 3411329 | christoph.scholz@iee.fraunhofer.de

Kognitive Energiesysteme

Automatisierte Erkennung von gefährdeten Vogelarten zum beschleunigten Ausbau von Windkraftanlagen

Grundsätzlich ist die Projektplanung von Windparks sehr aufwendig. Es müssen aufwendige Verträglichkeitsprüfungen in Bezug auf den Naturschutz, insbesondere für die Artengruppen Vögel und Fledermäuse, durchgeführt werden.

Im Vorhaben des im nächsten Jahr startenden BMUV Projekt »DeepBirdDetect« wird ein neuartiges, KI-gestütztes System zur automatisierten Detektion von windkraft-sensiblen und gefährdeten Arten entwickelt, welches die Genehmigungspraxis und somit den Ausbau von Windenergieanlagen deutlich beschleunigt. Ausgehend von einer automatisierten Erkennung und Klassifizierung anhand von Audiosignalen mittels Techniken, vor allem des Deep Learnings, erfolgt eine zeitliche/räumliche Erfassung des Vorkommens. In diesem Leuchtturm wurde der State-of-the-Art umgesetzt und getestet, sowie eine erste vergleichende Modellarchitektur entwickelt und evaluiert.

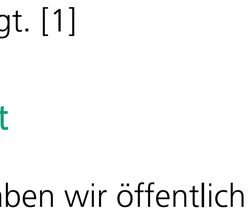
Problematik der Genehmigungsverfahren

Viele Windenergieprojekte können aufgrund von Klagen nicht in Betrieb gehen, wobei der Klagegrund bzgl. der Gefährdung besonders geschützter Vogel- bzw. Fledermausarten 48% beträgt. [1]

Daten und State-of-the-Art

Zum Training der Modelle haben wir öffentlich zugängliche Daten von xenocanto sowie Felddaten von einem Projektpartner verwendet. Im Bereich der Vogelstimmenerkennung stellt BirdNet aktuell den State-of-the-Art dar.

Tabelle 1: Wichtigste Eigenschaften der verwendeten Daten und von BirdNet

Daten	BirdNET (State-of-the-ART)		
Audioaufnahmen mit unterschiedlicher Qualität bzgl.:	**Erkennung**	**Einschränkung der Klassen**	**Genauigkeit**
▪ Annotation ▪ Hintergrundgeräusche ▪ Sounddetektion mittels 360-Grad-Mikrofon oder Richtmikrofon	Neben Vögeln auch andere Spezies und »Sound-Events«	Klassen gesamt: 3337	~60% – 80% Korrekte Erkennung
Neue Annotation »call« und »no call« hinzugefügt	Klassen gesamt: 3337	Nach Art: Manuell mithilfe von Spezies-Listen als Parameter	Schwankt stark mit der Qualität
Neue Daten mit »kollisionsgefährdeten Vogelarten« nach BNatSchG hinzugefügt - Anzahl: 15 - Erkannt in unserem Datensatz: 4 (Fischadler, Rotmilan, Schwarzmilan, Weißstorch)		Nach Zeit: Saisonale Unterschiede werden automatisch einbezogen	Optimierbar durch Parameter »Confidence« und »Sensitivität«

Datenvorverarbeitung

Zur Klassifizierung von Vogelstimmen haben wir ein ResNet-152 (siehe Abb. 2) auf den Mel-Spektogrammen der Audiodaten trainiert und evaluiert (Abb. 1.)

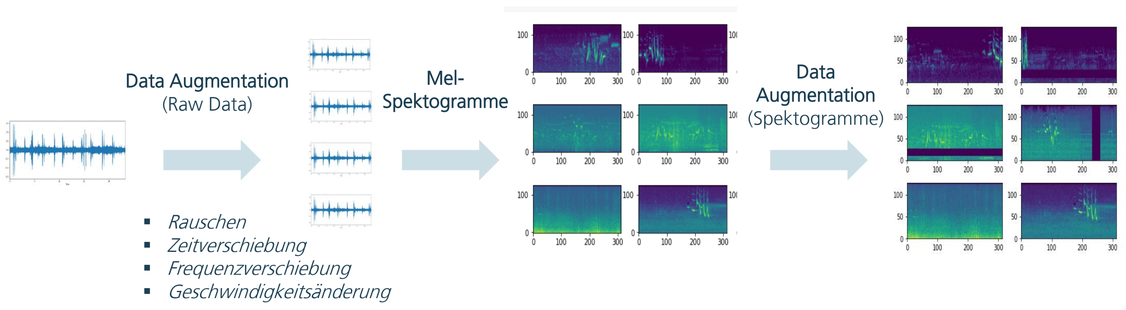


Abbildung 1: Zur Erstellung der Mel-Spektogramme wurden alle verfügbaren Audiodaten in 5-Sekunden-Blöcke (Chunks) unterteilt. Um weiterhin dem Problem des Overfitting zu entgegnen, wird Data Augmentation Techniken aus dem Audio-Bereich verwendet, um die Anzahl von Trainingsdaten zu erhöhen. Dabei werden sowohl Data Augmentation Techniken auf den Rohdaten als auch auf den Mel-Spektogrammen (Frequenz- und Zeitmaskierung) verwendet.

Modellarchitektur

Als Architektur des Modells zur Klassifizierung von Vogelstimmen verwenden wir ein Residual Network (ResNet) mit 152 Layern in der Version V2. Residual Networks wurden im Jahr 2015 vorgestellt und haben durch den Gewinn des ILSVRC-Wettbewerbs besondere Aufmerksamkeit erhalten. Mit ResNet war es durch die Einführung von "Skip-Connections" möglich, besonders tiefe Neuronale Netze zu trainieren. Das Gewinnernetz des ILSVRC-Wettbewerbs benutzte hier ein tiefes "Convolutional Neural Network" mit 152 Schichten. Im Jahr 2016 wurde die Architektur der ResNets in der Version V2 angepasst, indem Batch-Normalisierung und ReLU-Aktivierung auf die Eingabe angewendet wird, bevor die Faltungsoperation erfolgt. In der Version 1 wird zunächst die Faltungsoperation auf die Eingabe angewendet.

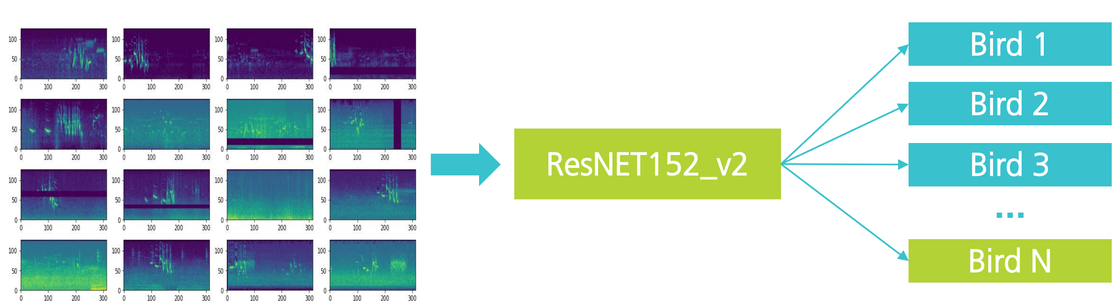


Abbildung 2: Modellarchitektur

Ergebnisse

Insgesamt haben wir für drei Szenarien Modelle trainiert und getestet.

- **Szenario 1 (Klassen: N=131)** umfasst qualitativ hochwertige Audiodaten (Score >= 4.0) mit einer ausreichend großen Menge an Trainingsdaten (pro Klasse > 100 Audiodaten).
- **In Szenario 2 (N=364)** nutzen wir alle Trainingsdaten, wobei mindestens 30 Audiodaten für eine Klasse verfügbar sind.
- **Szenario 3 (N=2)** betrachtet die Unterscheidung des Erkennens von Rufen (Calls) und keinen Rufen (no Calls).

Tabelle 2: Ergebnisse

Modell	Szenario 1	Szenario 2	Szenario 3
ResNet-152	65 %	55 %	95 %
BirdNet	69 %	59 %	-
CNN	27 %	-	-

Gefördert durch:

HESSEN
Hessisches Ministerium für Wissenschaft und Kunst

1 Quelle: Fachagentur Windenergie an Land »Hemmnisse beim Ausbau der Windenergie in Deutschland«, 2019

SPOTLIGHTS

Fraunhofer IEE