

Vogelstimmenerkennung mittels Gabor-Frames

Rhein-Ruhr-Workshop 2020

Sven Heuer

Philipps-Universität Marburg

01. Februar 2020

Vogelstimmen- erkennung mittels Gabor-Frames

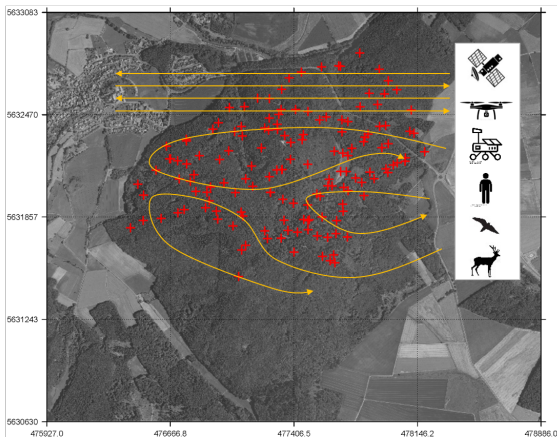
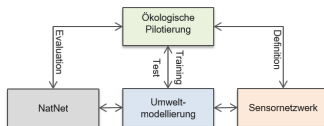
Sven Heuer

Motivation

Verfahren zur Vogelstimmen- erkennung

Gabor-Transformation
Preprocessing
Deep Learning

Kompressions- verfahren mit Gabor-Frames



Motivation

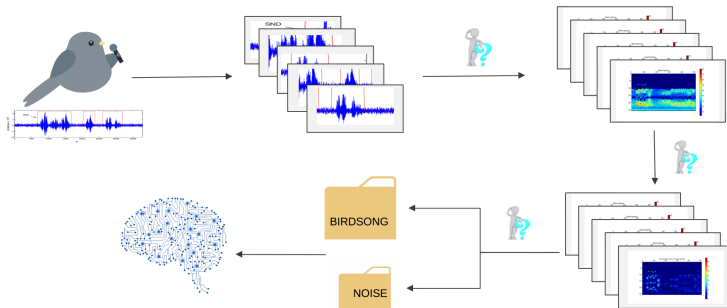
Verfahren zur Vogelstimmener- kennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions- verfahren mit Gabor-Frames



Definition

Sei $0 \neq g \in L_2(\mathbb{R}^d)$ eine sogenannte *Fensterfunktion*. Die STFT (short-time Fourier transform) eines Signals $f \in L_2(\mathbb{R}^d)$ bezüglich g ist gegeben durch

$$V_g f(x, \omega) = \int_{\mathbb{R}^d} f(t) \overline{g(t-x)} e^{-2\pi i t \cdot \omega} dt, \quad x, \omega \in \mathbb{R}^d.$$

Die *Gabor-Transformation* von f bezüglich g entspricht der Abtastung von $V_g f(x, \omega)$ auf einem Gitter $\alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$.

Vogelstimmen-
erkennung mittels
Gabor-Frames

Sven Heuer

Motivation

Verfahren zur
Vogelstimmen-
erkennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames

Definition

Sei $0 \neq g \in L_2(\mathbb{R}^d)$ eine sogenannte *Fensterfunktion*. Die STFT (short-time Fourier transform) eines Signals $f \in L_2(\mathbb{R}^d)$ bezüglich g ist gegeben durch

$$V_g f(x, \omega) = \int_{\mathbb{R}^d} f(t) \overline{g(t-x)} e^{-2\pi i t \cdot \omega} dt, \quad x, \omega \in \mathbb{R}^d.$$

Die *Gabor-Transformation* von f bezüglich g entspricht der Abtastung von $V_g f(x, \omega)$ auf einem Gitter $\alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$.

Bemerkung: Mit

$$\begin{aligned} M_\omega g(t) &= e^{2\pi i \omega \cdot t} g(t), \\ g^*(t) &= \overline{g(-t)} \end{aligned}$$

gilt

$$V_g f(x, \omega) = e^{-2\pi i x \cdot \omega} (f * M_\omega g^*)(x).$$

Vogelstimmen-
erkennung mittels
Gabor-Frames

Sven Heuer

Motivation

Verfahren zur
Vogelstimmen-
erkennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames

Spektrogramm Dorngrasmücke

Vogelstimmen-
erkennung mittels
Gabor-Frames

Sven Heuer

Motivation

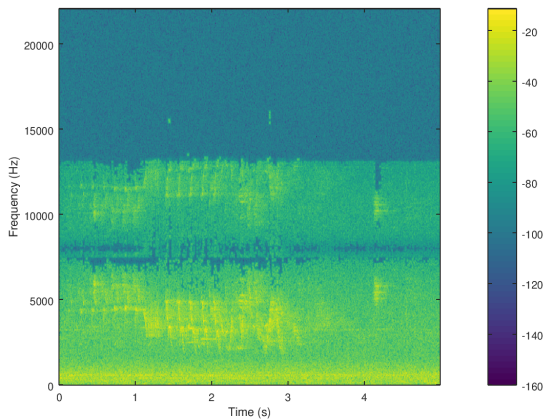
Verfahren zur
Vogelstimmener-
kennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames



$$SPEC_g f(x, \omega) = \log_{10}(|V_g f(x, \omega)|)$$

Motivation

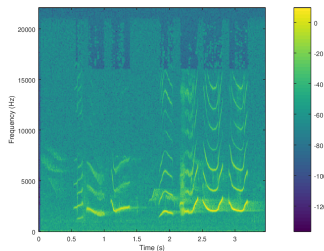
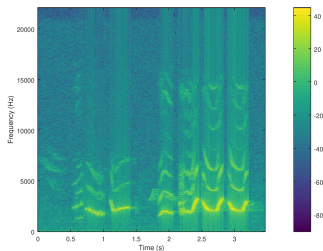
Verfahren zur Vogelstimmen- erkennung

Gabor-Transformation

Preprocessing

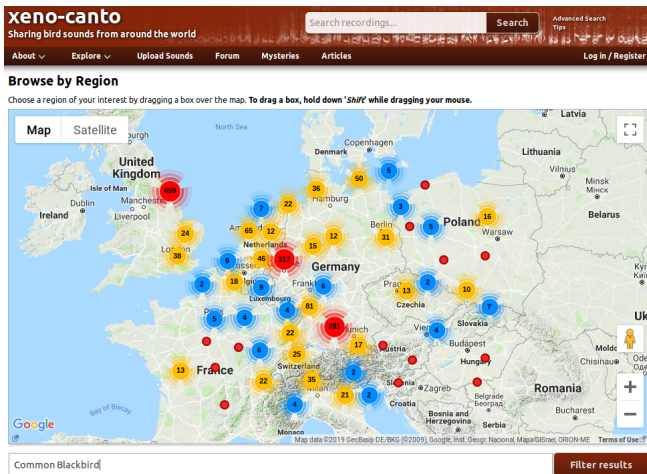
Deep Learning

Kompressions-
verfahren mit
Gabor-Frames



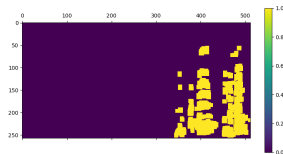
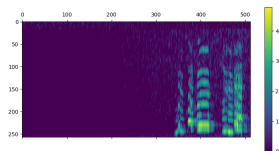
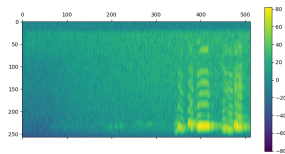
Links: „Rechteck“-Fenster: $g = \chi_{[-1,1]}$

Rechts: Gauß-Fenster: $g_a(x) = e^{-\pi x^2/a}$, $a > 0$

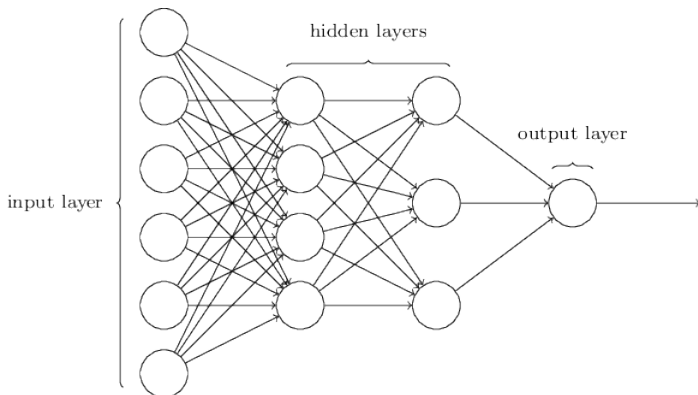


2040 Recordings from 1 species (only the first 2000 recordings are shown on the map) found in the selected region

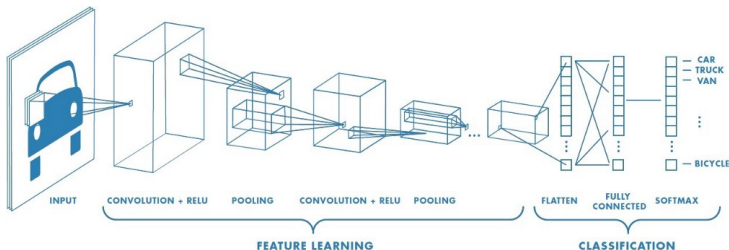
- 74 Vogelarten
- 1276 Audiodateien (MP3-Format)
- Insgesamt ca. 26 Stunden Aufnahmen



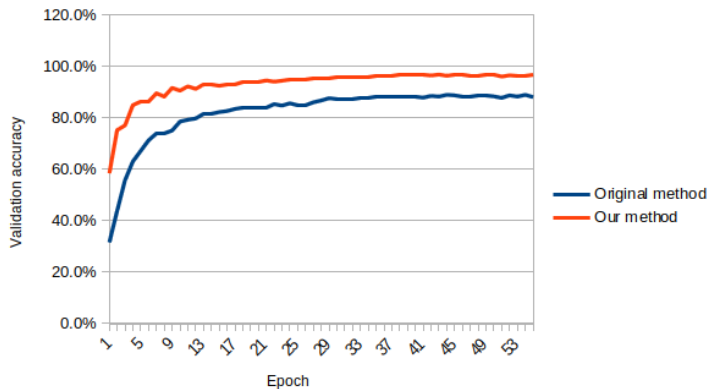
1. Betrachte
5-Sekunden-Ausschnitt
2. Berechne Spektrogramm
(erhöhte Zeitauflösung)
3. Entrausche Spektrogramm
zeilenweise, basierend auf
Quantil
4. Eliminiere isolierte Pixel



- Gewichtete Summen, Gewichte werden erlernt
- Sehr(!) viele Verknüpfungen
- Besser: Eigenschaften/Struktur von Bildern nutzen



- Faltungen reduzieren Dimension
- Faltungskerne werden erlernt
- Tiefere Strukturen möglich



Erinnerung:

$$|V_g f(x, \omega)| = |(f * M_\omega g^*)(x)|.$$

Motivation

Verfahren zur
Vogelstimmener-
kennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames

Erinnerung:

$$|V_g f(x, \omega)| = |(f * M_\omega g^*)(x)|.$$

- Gabor-Transformation direkt in CNN integrieren
 - + „Flexiblere“ Strukturen
 - + Schnellere Berechnungen(?)
 - Data Augmentation schwieriger
- Zeitkomponente in Verfahren einfließen lassen
- Datenkompression

Motivation

Verfahren zur
Vogelstimmen-
erkennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames

Definition

Betrachte das Gitter $\Lambda = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$ und eine Fensterfunktion $g \in L_2(\mathbb{R}^d)$. Die Menge

$$\{g_\lambda = (T_{\alpha k} M_{\beta n} g) \mid \lambda = (\alpha k, \beta n) \in \Lambda\}$$

heißt *Gabor-Frame*, falls $0 < A \leq B < \infty$ existieren, sodass

$$A \|f\|^2 \leq \sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|^2 \leq B \|f\|^2$$

für alle $f \in L_2(\mathbb{R}^d)$ gilt.

Lemma

Sei $\{g_\lambda \mid \lambda \in \Lambda\}$ ein Gabor-Frame. Dann existiert ein duales Fenster $\tilde{g}$, sodass $\{\tilde{g}_\lambda \mid \lambda \in \Lambda\}$ ein Frame ist und jedes $f \in L_2(\mathbb{R}^d)$ geschrieben werden kann als

$$f = \sum_{\lambda \in \Lambda} \langle f, \tilde{g}_\lambda \rangle g_\lambda$$

beziehungsweise

$$f = \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle \tilde{g}_\lambda.$$

Definition

Sei $0 \neq g \in \mathcal{S}(\mathbb{R}^d)$ eine Fensterfunktion im Schwartz-Raum. Der *Modulationsraum* $\mathcal{M}_p(\mathbb{R}^d)$, $0 < p < \infty$, ist gegeben durch

$$\mathcal{M}_p(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \rightarrow \mathbb{C} \mid V_g f \in L_p(\mathbb{R}^{2d}) \right\}$$

mit der Norm

$$\|f\|_{\mathcal{M}_p(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_g f(x, \omega)|^p dx d\omega \right)^{1/p}.$$

Lemma (Rauhut, 2005)

Sei $0 < p < \infty$, $g \in \mathcal{M}_p(\mathbb{R}^d)$ und $\Lambda = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$ so, dass $\{g_\lambda \mid \lambda \in \Lambda\}$ einen Gabor-Frame bildet. Dann existieren Konstanten $C_1, C_2, \tilde{C}_1, \tilde{C}_2 > 0$, sodass

$$C_1 \|f\|_{\mathcal{M}_p(\mathbb{R}^d)} \leq \|(\langle f, g_\lambda \rangle)_{\lambda \in \Lambda}\|_{\ell_p} \leq C_2 \|f\|_{\mathcal{M}_p(\mathbb{R}^d)},$$

$$\tilde{C}_1 \|f\|_{\mathcal{M}_p(\mathbb{R}^d)} \leq \|(\langle f, \tilde{g}_\lambda \rangle)_{\lambda \in \Lambda}\|_{\ell_p} \leq \tilde{C}_2 \|f\|_{\mathcal{M}_p(\mathbb{R}^d)}$$

für alle $f \in \mathcal{M}_p(\mathbb{R}^d)$ gelten.

Voraussetzungen:

- $p \in (0, 2)$
- $g \in \mathcal{M}_p(\mathbb{R}^d)$, Λ wie zuvor
- $f \in \mathcal{M}_p(\mathbb{R}^d)$

Satz

Für $\mu > 0$, setze $I_\mu = \{\lambda \in \Lambda \mid |\langle f, \tilde{g}_\lambda \rangle| \geq \mu\}$ und

$$f_\mu = \sum_{\lambda \in I_\mu} \langle f, \tilde{g}_\lambda \rangle g_\lambda.$$

Dann gilt

$$\|f - f_\mu\|_{L_2}^2 \leq \tilde{C}_2^2 \|f\|_{\mathcal{M}_p(\mathbb{R}^d)}^2 N^{1-2/p}$$

mit $N = \#I_\mu$.

Vogelstimmen-
erkennung mittels
Gabor-Frames

Sven Heuer

Motivation

Verfahren zur
Vogelstimmener-
kennung

Gabor-Transformation

Preprocessing

Deep Learning

Kompressions-
verfahren mit
Gabor-Frames



M. A. Nielsen.

Neural Networks and Deep Learning. Determination Press. 2015. (Neuronales Netz)



Online-Quellen.

Singvogel: <https://pixabay.com/illustrations/sing-bird-songbird-microphone-1322180>

CNN: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

 S. Heuer, P. Tafo, S. Dahlke und H. Holzmann.
New aspects in birdsong recognition utilizing the gabor
transform. In: Proceedings of the 23rd International
Congress on Acoustics. Aachen, 2019.

 K. Gröchenig.
Foundations of Time-Frequency Analysis.
Birkhäuser/Springer, New York. 2001.

 H. Rauhut.
Coorbit Space Theory for Quasi-Banach Spaces.
arXiv:math/0507466v1. 2005.