

# Lightweight Multi-Branch Network for Animal Individual Re-Identification

---

22.09.2021

Bertan Karacora  
Institut für Informatik, Universität Bonn  
[bertan.karacora@uni-bonn.de](mailto:bertan.karacora@uni-bonn.de)

# Inhalt

---

1. **Einleitung**
2. Daten und Annotation
3. Methoden
4. Ergebnisse und Evaluation
5. Zusammenfassung

# Animal Individual Re-Identification

Das Wiedererkennen eines individuellen Tieres in einer Galerie aus Bildern, gegeben ein einzelnes Beispiel-Bild des Individuums.<sup>[6]</sup>

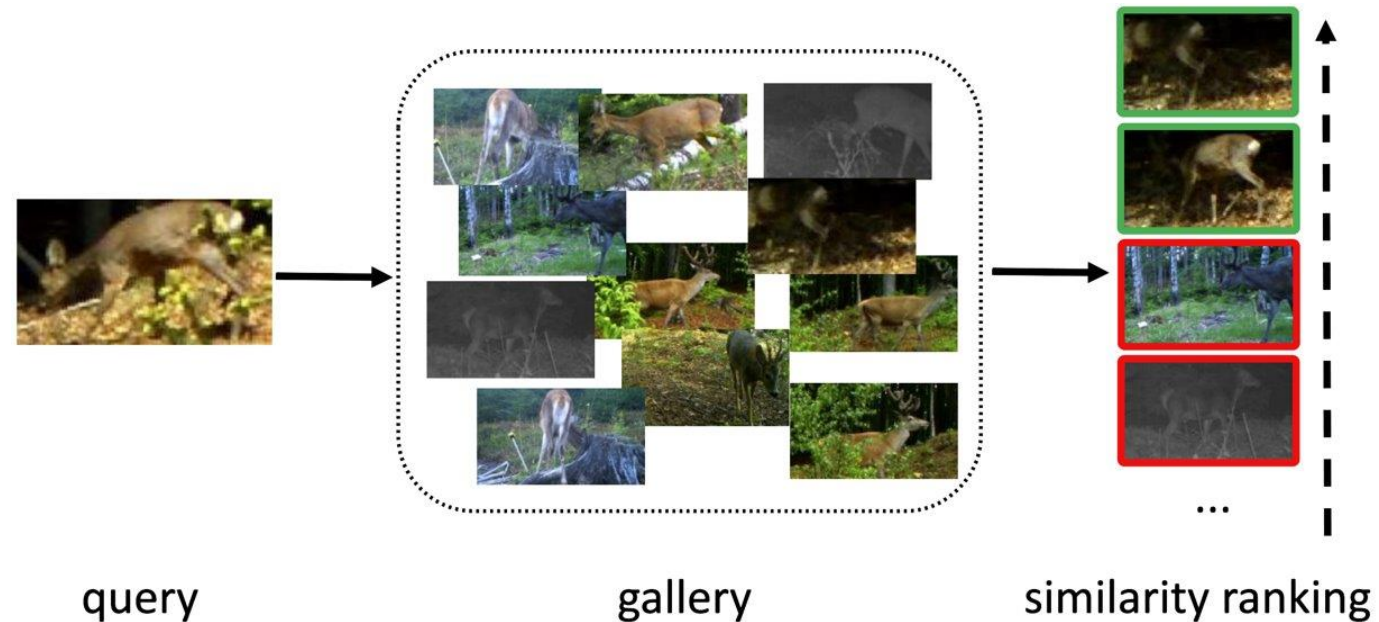


Abb. 1: Animal Individual Re-Identification mithilfe eines Ähnlichkeits-Maßes.

# Verwendung

---

In biologischer/ökologischer Forschung:[10, 11, 12]

- Beobachtung des Tierverhaltens
- Verwendung mehrerer Kameras über beliebige Zeiträume (*camera traps*)
- Soziale Strukturen
- Schätzungen zu Vorkommen und Vielfalt
- Nachhaltiger Umgang mit Wildtier-Beständen

# Bisherige Ansätze

---

## Körperliche Verfahren<sub>[9, 12]</sub>

- Tagging der Tier-Körper
- Implantate
- Microchips
- Gentechnik
- Individuelle Markierungen oder Verletzungen

## Computer Vision

- Person Re-Identification als klassische Problemstellung der Computer Vision
- Freytag et al.: Gesichtserkennung von Schimpansen mit dem AlexNet<sub>[4]</sub>
- Brust et al.: Detektion und Wiedererkennung von Gorillas<sub>[1]</sub>
- Schneider et al.: Animal Re-Identification durch Ähnlichkeitsmaße<sub>[11]</sub>

# Ziel dieser Arbeit

---

- Umsetzung eines Wiedererkennungssystems auf Videodaten von Wildtieren
- Anwendung, Analyse und Adaption des Lightweight Multi-Branch Networks (Herzog et al.<sup>[6]</sup>)
- Beobachtung, Visualisierung und Evaluation der Performanz unter Variation von Trainings- und Testparametern sowie der Netzwerk-Architektur

# Inhalt

---

1. Einleitung
- 2. Daten und Annotation**
3. Methoden
4. Ergebnisse und Evaluation
5. Zusammenfassung

# Daten und Annotation

---

- BayWald und Wildpark
  - Rothirsch und Europäisches Reh (*Cervus elaphus* und *Capreolus capreolus*)
  - Geringe Anzahl an Individuen und Mehrfach-Sichtungen
  - Ähnliche Aufnahmebedingungen der Videos
  - Background Bias durch Video-Daten
- Annotation
  - Bounding Boxes oder Segmentierungsmasken
  - MiVOS<sub>[2]</sub> und VGG Image Annotator<sub>[3]</sub>



# BayWald

---

- 25 Videos, 8 oder 15 FPS, 6292 Bilder
- 20 Instanzen, 1 Instanz/Bild, 3 Instanzen mit mehrfacher Sichtung
- Besonderheiten und Beispiele:

Umgebung



Nachtaufnahmen



Lichtverhältnisse

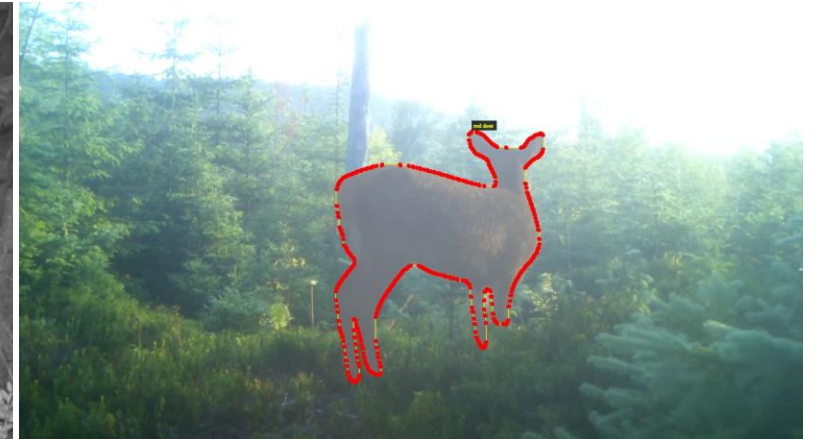


Abb. 2: Beispielbilder mit Polygon-Umriss aus BayWald

# Wildpark

- 4 Videos, 30 FPS, 3600 Bilder
- 34 Instanzen, 5-12 Instanzen/Bild, bis zu 21 verschiedene Individuen, mehrfache Sichtungen unbekannt
- Besonderheiten und Beispiele:

Gegenseitige  
Überdeckung  
(insbes. der  
Bounding Boxes)



Niedrige Auflösung



Abb. 3: Beispielbilder mit Polygon-Umriss aus Wildpark

# MiVOS

---

- Interaktives Tool für die Annotation von Videos
- Iterativ-alternierende Nutzung von 3 Modulen:
  - Interaction-to-mask
  - Propagation
  - Fusion
- Limitierung in Anzahl zu annotierender Objekte und Auflösung
- Einschränkungen durch Hardware und Beschaffenheit der Daten



# VGG Image Annotator

- Manuelle Annotation einzelner Bilder mithilfe von geometrischen Formen
- Copy-Funktion für Videoframes ohne Tierbewegung



Abb. 4: Vergleich von Segmentierungs-Ergebnissen nach Transformation zu Polygonen. Links: MiVOS. Rechts: VGG Image Annotator

# Inhalt

---

1. Einleitung
2. Daten und Annotation
- 3. Methoden**
4. Ergebnisse und Evaluation
5. Zusammenfassung

# Netzwerk-Architektur

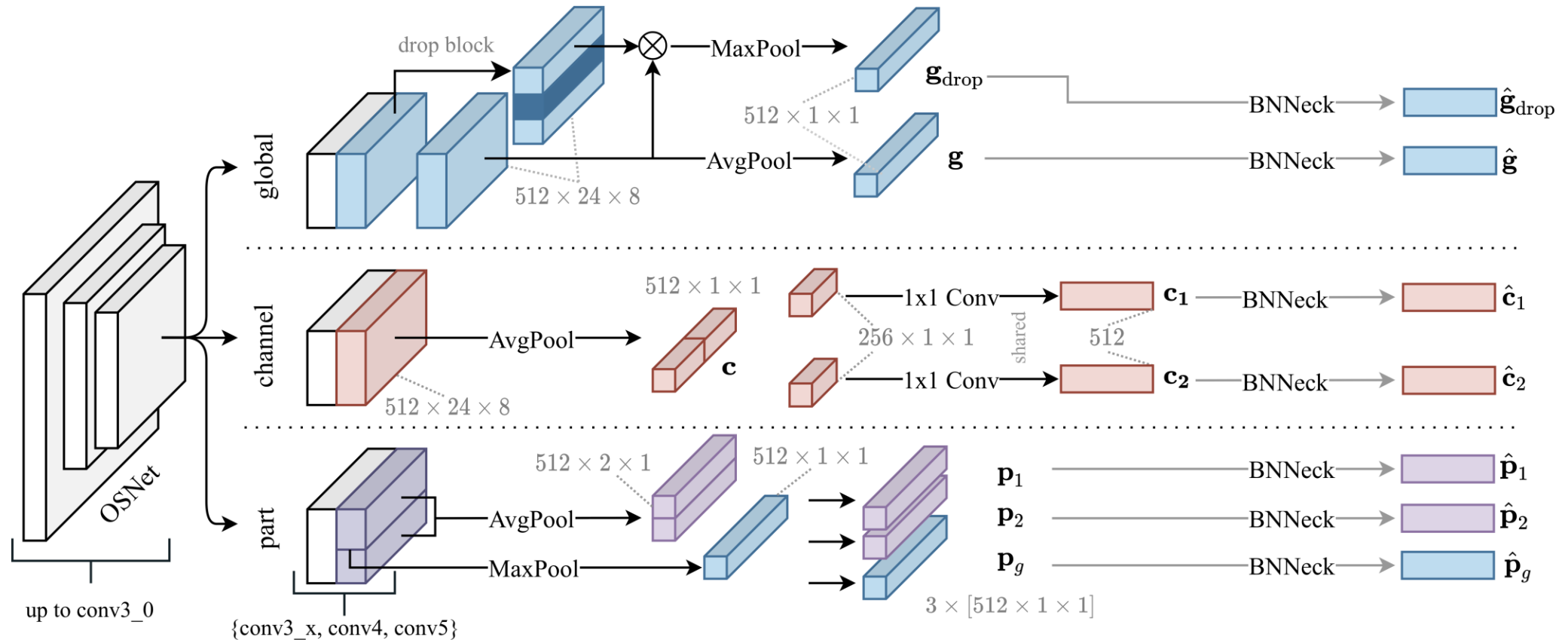


Abb. 4: Architektur des Lightweight Multi-Branch Networks. Quelle: [6].

# Optimierung des Trainings

---

- Kombination von Cross-Entropy und Multi-Similarity Loss<sub>[13]</sub>
- Random Erasing Augmentation
- Label Smoothing zur Verringerung von Overfitting
- Warmup Cosine Annealing Schedule für die Learning Rate
- Verringerte Learning Rate wegen niedriger Instanzen-Anzahl und Ähnlichkeit der Videos

# Verarbeitung der Video-Daten

---

- Video-Daten unterscheiden sich stark von ursprünglich verwendeten Personen-Datensätzen Market1501 und CUHK03
- Pre-processing der Video-Daten:
  - 1) Laden der Annotations-Dateien und Videoframes.
  - 2) Zuschneiden der Bilder zu Bounding Boxes oder Segmentierungs-Masken mithilfe der annotierten Polygone.
  - 3) Split der Bilder in Trainings-, Query- und Gallery-Set.
  - 4) Relabeling der Individuen für das Training.
  - 5) Ausgabe der Daten auf Image-level
- Implementation mit Pytorch<sub>[8]</sub> und Torchreid<sub>[15, 16]</sub>



# Zuschneiden der Bilder

---

- Detektion als Voraussetzung für Re-Identification
- Verringerung des Background Bias
- Unterscheidung der Instanzen in Wildpark



Abb. 5: Zuschneiden eines Bildes aus BayWald auf Bounding Box und Segmentierungsmaske.

# Inhalt

---

1. Einleitung
2. Daten und Annotation
3. Methoden
- 4. Ergebnisse und Evaluation**
5. Zusammenfassung

# Experimentelle Durchführung

---

- BayWald: 30 Epochen, 3 x 10 Batchsize, 320 x 180 Auflösung
- Wildpark: 20 Epochen, 3 x 10 Batchsize, 240 x 180 Auflösung
- Evaluationsmetriken von Herzog et al.:<sup>[6]</sup>
  - Rank-k Accuracy
  - Mean Average Precision (mAP)
- Evaluation auf Wildpark entweder nur innerhalb von Videos oder unter Vernachlässigung von Mehrfach-Sichtungen

# Trainingsparameter und Performanz

Methode	BayWald	Wildpark
Baseline	0,9873	0,9516
FD	0,9949	0,9467
QGT	0,9833	0,9350
FD + QGT	0,9938	0,9364
FD + QGT + LS + WCA	0,9885	0,9320
FD + QGT + MA	0,9994	0,9225
FD + QGT + LS + WCA + MA	0,9995	0,9225
TRA + FD + QGT	-	0,9241

FD: Frame Dropping

QGT: Query-Gallery-Trennung

LS: Label Smoothing

WCA: Warmup Cosine  
Annealing

MA: Training auf Masken statt  
Bounding Boxes

TRA: Transfer des auf BayWald  
trainiertes Netzwerks auf  
Wildpark

Tabelle 1: mAP pro Methode auf BayWald und Wildpark.

# mAP während des Trainings

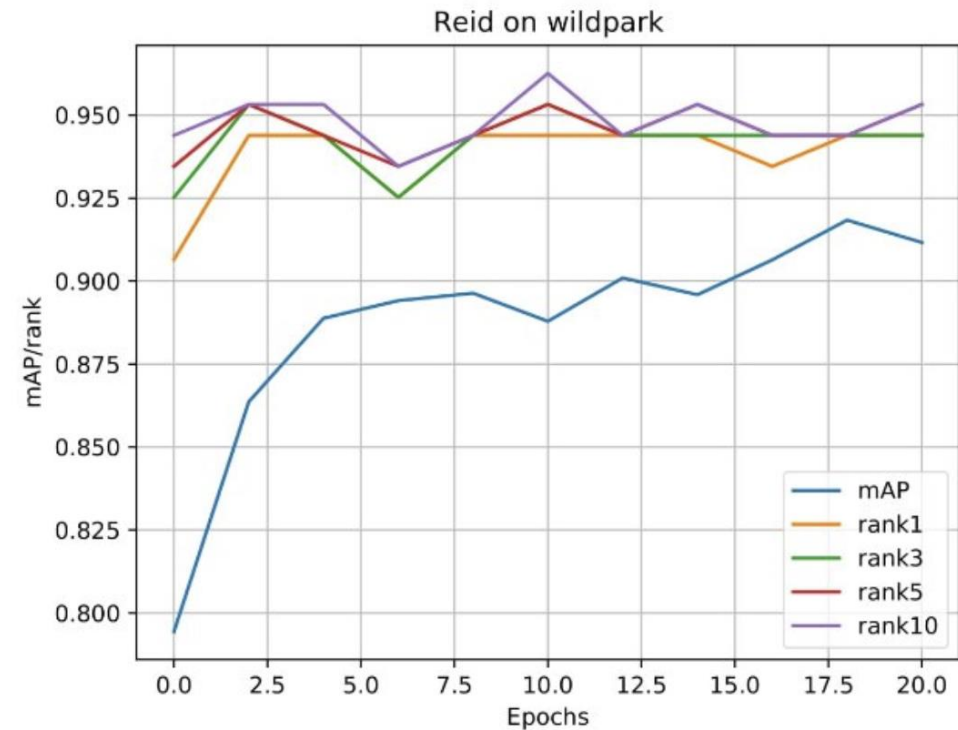
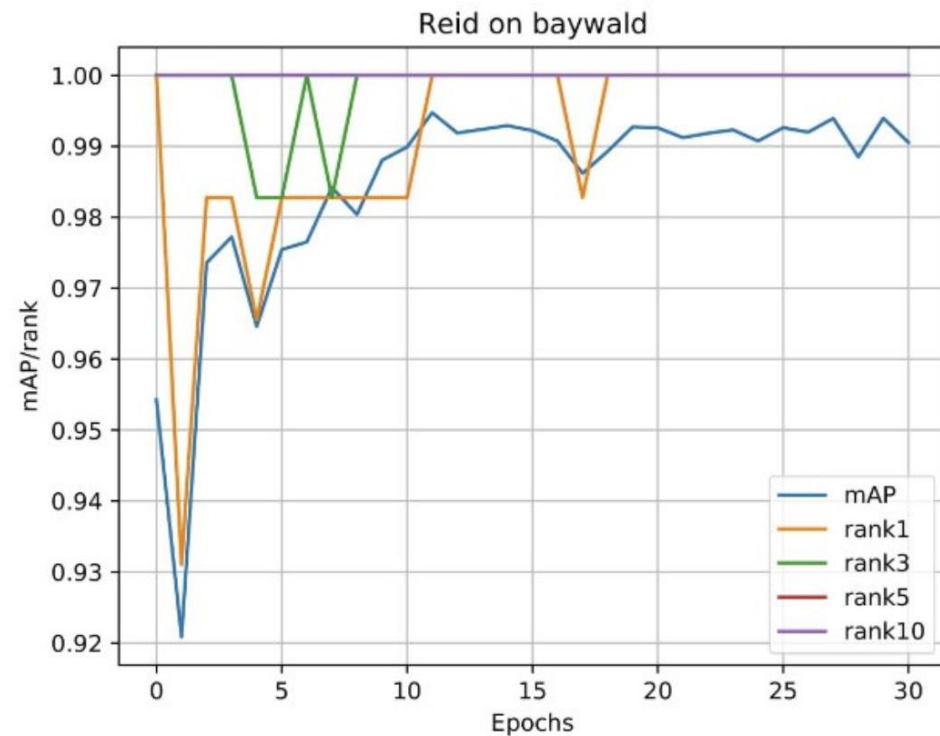


Abb. 6: mAP/Epoche auf BayWald (links) und Wildpark (rechts). Konfiguration: FD + QGT.

# Activation Maps

---

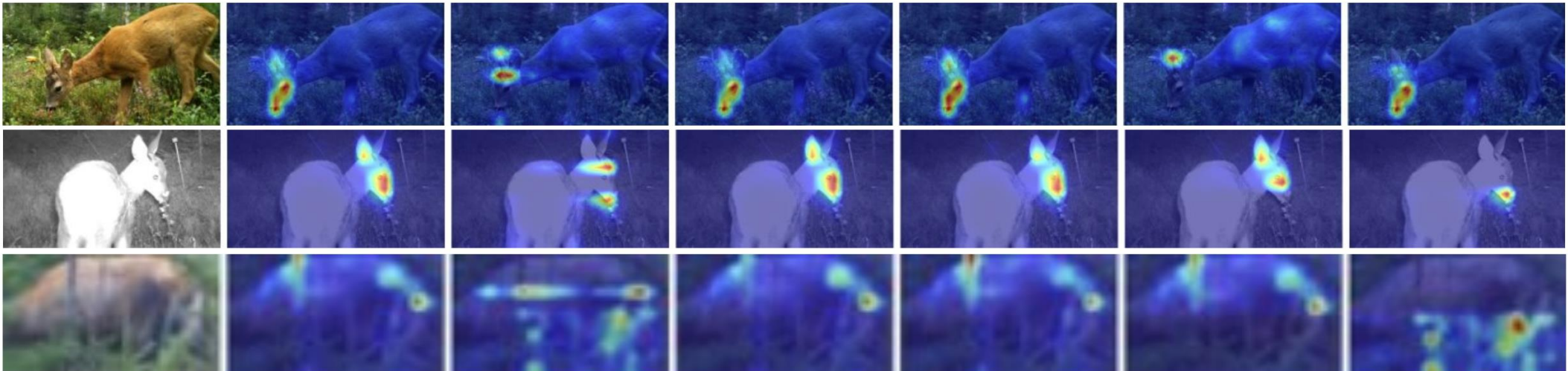


Abb. 7: Activation Maps. Von links nach rechts: Original-Bild, globale Features, globale Features mit Drop Block, erste Hälfte der Channel-Branches, zweite Hälfte, oberer Teil des Part-Branches, unterer Teil.



# Similarity Rankings

---



Abb. 8: Ranking der ähnlichsten Gallery-Bilder (rechts) bezüglich eines Query-Bildes (links).

# Netzwerk-Komplexität

---

Dataset	Parameter	Flops
BayWald	$5,07 \cdot 10^6$	$2,80 \cdot 10^9$
Wildpark	$6,70 \cdot 10^6$	$2,10 \cdot 10^9$
Market1501	$9,35 \cdot 10^6$	$2,43 \cdot 10^9$

Tabelle 2: Schätzung der Netzwerk-Komplexität auf BayWald, Wildpark und Market1501. Parameter: Gewichte and Parameter zur Beschreibung des Netzwerks. Flops: Floating Point Operations Per Second.



# Frame Dropping

---

- Ausdünnen der Video-Frames zur Verringerung von Background Bias und Redundanz
- Überspringen von ähnlichen und unähnlichen Frames gleichermaßen für ein gegebenes Query-Bild
- Gemäßigte Anwendung hinsichtlich Datenmenge, Videolänge, Framerate und Tierbewegung
- Verringerung von Background Bias nur bei ausreichender Bewegung des Tieres
- Rank-k Accuracy ohne Frame Dropping oft konstant

# Auswirkung des Frame Droppings

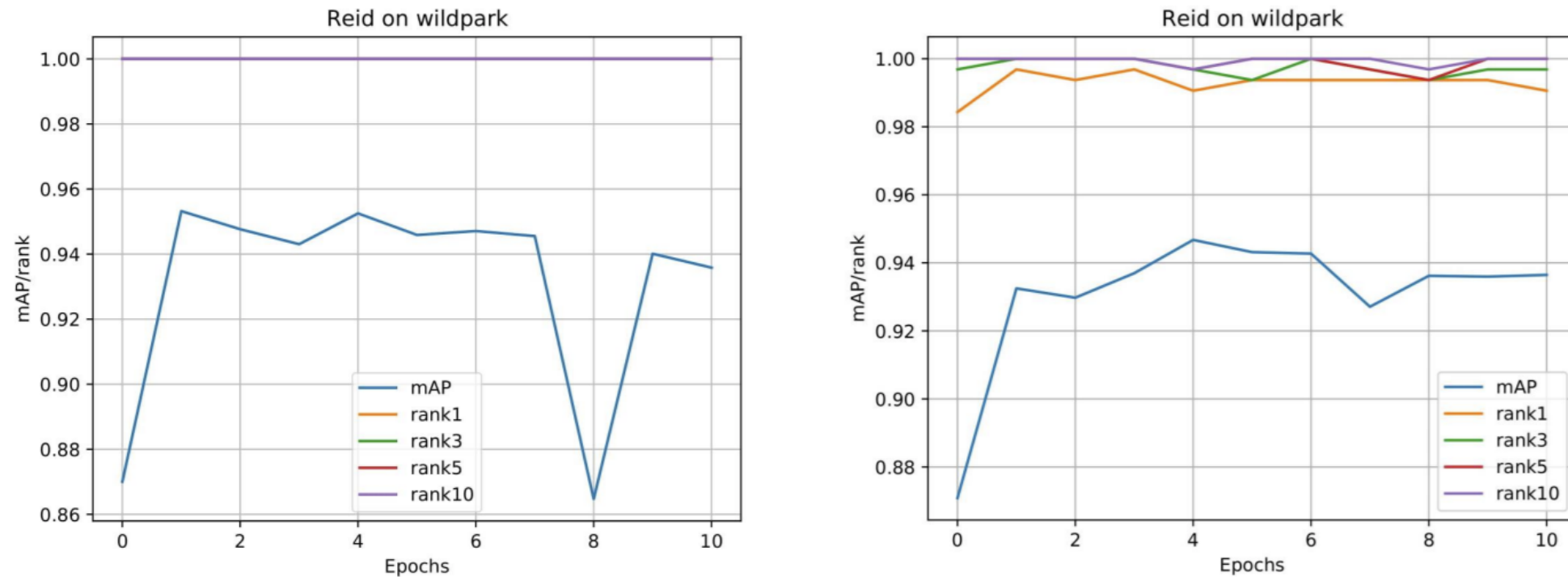


Abb. 9: mAP/Epoche auf Wildpark. Links: Baseline. Rechts: FD.

# Query-Gallery-Trennung

---

- Zeitliche Trennung von Query- und Gallery-Set innerhalb eines Videos
- Verringerung der Ähnlichkeit zwischen Query und Gallery
- Verringerung von Background Bias bei Tierbewegung in ausgenommener Zeitspanne
- Geringere mAP des pretrained Netzes (0. Epoche), aber Wiedererkennung bei weiterem Training wenig eingeschränkt

# Variation der Architektur

Architektur	BayWald	Wildpark
OS + G	0.9836	0.8944
OS + C	0.9958	0.9322
OS + P	0.9928	0.9273
OS + G + C + P	0.9995	0.9320
RES + G + C + P	0.9957	0.9117

OS: OSNet Backbone

RES: ResNet50 Backbone

G: Global Branch

C: Channel Branch

P: Part Branch

Tabelle 3: mAP pro Netzwerk-Architektur auf BayWald und Wildpark.

# Masken statt Bounding Boxes

---

- Kein Background Bias möglich
- Kontext geht verloren
- Keine Verbesserung in Activation Maps erkennbar, da die Activation ohnehin nur auf dem Tierkörper zu sehen ist
- Erhöhte Performanz bei BayWald, leichte Verringerung bei Wildpark

# Zusammenfassung

---

- Hohe Performanz des Lightweight Multi-Branch Networks auf video-basierten Wildtier-Datensätzen
- Höchste Performanz auf BayWald mit Multi-Branch Struktur und Kombination aller Trainingstechniken
- Training auf Masken kann zu weiterer Verbesserung führen
- Keine Veränderung der Netz-Komplexität
- Verbesserte Evaluation möglich mithilfe von Frame Dropping und Query-Gallery-Trennung

# Referenzen

1. Brust, C. ; Burghardt, T. ; Groenenberg, M. ; Käding, Christoph ; Kühl, H. ; Manguette, Marie L. ; Denzler, Joachim: Towards Automated Visual Monitoring of Individual Gorillas in the Wild. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) (2017), S. 2820–2830
2. Cheng, Ho K. ; Tai, Yu-Wing ; Tang, Chi-Keung: Modular Interactive Video Object Segmentation: Interaction-to-Mask, Propagation and Difference-Aware Fusion. In: CoRR abs/2103.07941 (2021). – URL <https://arxiv.org/abs/2103.07941>
3. Dutta, Abhishek ; Zisserman, Andrew: The VIA Annotation Software for Images, Audio and Video. In: Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA : Association for Computing Machinery, 2019 (MM'19), S. 2276–2279. – URL <https://doi.org/10.1145/3343031.3350535>. – ISBN 9781450368896
4. Freytag, Alexander ; Rodner, Erik ; Simon, Marcel ; Loos, Alexander ; Kühl, Hjalmar S. ; Denzler, Joachim: Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Identities and Attributes of Primates. (2016), S. 51–63. ISBN 978-3-319-45886-1
5. He, Kai ming ; Zhang, Xiangyu ; Ren, Shaoqing ; Sun, Jian: Deep Residual Learning for Image Recognition. In: CoRR abs/1512.03385 (2015). – URL <http://arxiv.org/abs/1512.03385>
6. Herzog, Fabian ; Ji, Xunbo ; Teepe, Torben ; Hörmann, Stefan ; Gilg, Johannes ; Rigoll, Gerhard: Lightweight Multi-Branch Network for Person Re-Identification. In: CoRR abs/2101.10774 (2021). – URL <https://arxiv.org/abs/2101.10774>
7. Li, Wei ; Zhao, Rui ; Xiao, Tong ; Wang, Xiaogang: DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, S. 152–159
8. Paszke, Adam ; Gross, Sam ; Massa, Francisco ; Lerer, Adam ; Bradbury, James ; Chanan, Gregory ; Killeen, Trevor ; Lin, Zeming ; Gilmershein, Natalia ; Antiga, Luca ; Desmaison, Alban ; Köpf, Andreas ; Yang, Edward ; DeVito, Zach ; Raison, Martin ; Tejani, Alykhan ; Chilamkurthy, Sasank ; Steiner, Benoit ; Fang, Lu ; Bai, Junjie ; Chintala, Soumith: PyTorch: An Imperative Style, 13 High-Performance Deep Learning Library. In: CoRR abs/1912.01703 (2019). – URL <http://arxiv.org/abs/1912.01703>
9. Phillott, Andrea ; Skerratt, Lee ; McDonald, Keith ; Lemckert, Frank ; Hines, Harry ; Clarke, John ; Alford, Ross ; Speare, Rick: Toe-Clipping as an Acceptable Method of Identifying Individual Anurans in Mark Recapture Studies. In: Herpetological Review 38 (2007), 09, S. 305–308
10. Schindler, Frank ; Steinhage, Volker: Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. In: Ecological Informatics 61 (2021), S. 101215. – URL <https://www.sciencedirect.com/science/article/pii/S1574954121000066>. – ISSN 1574-9541
11. Schneider, Stefan ; Taylor, Graham W. ; Linquist, Stefan S. ; Kremer, Stefan C.: Similarity Learning Networks for Animal Individual Re-Identification - Beyond the Capabilities of a Human Observer. In: CoRR abs/1902.09324 (2019). – URL <http://arxiv.org/abs/1902.09324>
12. Vidal, Maxime ; Wolf, Nathan ; Rosenberg, Beth ; Harris, Bradley P. ; Mathis, Alexander: Perspectives on Individual Animal Identification from Biology and Computer Vision. In: Integrative and Comparative Biology (2021), 05. – URL <https://doi.org/10.1093/icb/icab107>. – ISSN 1540-7063
13. Wang, Xun ; Han, Xintong ; Huang, Weiling ; Dong, Dengke ; Scott, Matthew R.: Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning. In: CoRR abs/1904.06627 (2019). – URL <http://arxiv.org/abs/1904.06627>
14. Zheng, Liang ; Shen, Liye ; Tian, Lu ; Wang, Shengjin ; Wang, Jingdong ; Tian, Qi: Scalable Person Re-identification: A Benchmark. In: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, S. 1116–1124
15. Zhou, Kaiyang ; Xiang, Tao: TorchReID: A Library for Deep Learning Person Re-Identification in Pytorch. In: CoRR abs/1910.10093 (2019). – URL <http://arxiv.org/abs/1910.10093>
16. Zhou, Kaiyang ; Yang, Yongxin ; Cavallaro, Andrea ; Xiang, Tao: Omni-Scale Feature Learning for Person Re-Identification. In: CoRR abs/1905.00953 (2019). – URL <http://arxiv.org/abs/1905.00953>