Master Thesis Proposal Nr.

Topic: Development of a Modular Software Framework for video-based Animal Re-Identification: Integration and Evaluation of Multiple Datasets and Models

Automated animal re-identification (Re-ID) plays a crucial role in wildlife monitoring, enabling researchers to track individual animals over time and study behavioral patterns. Traditional methods often rely on image-based approaches that disregard valuable temporal dynamics present in video sequences. This limitation hinders the ability to distinguish individuals within a species, particularly when dealing with night-time footage or non-patterned animals. Our project seeks to enhance video-based animal Re-ID by leveraging advanced temporal modeling techniques and pose estimation, integrating state-of-the-art deep learning architectures to improve identification accuracy. We present a Re-ID framework that leverages multiple video-based architectures trained on the PolarBearVid dataset [1], the BatVidID [2] and a novel GrizzlyVidID Dataset provided by the BearID project [1].

Our approach builds upon the Multi-direction and Multi-scale Pyramid Transformer for Video-based Pedestrian Retrieval (Pit) [3] leveraging the vision transformer backbone with different fusion strategies, which effectively capture fine-grained details, while the attention-based mechanism enhances robustness against occlusions. Additionally, we modify the Global-Local Temporal Representations for Video Person Re-Identification framework (GLTR)[4] by replacing the ResNet backbone with the MIEW-ID model [5], specifically designed for multi-species animal Re-ID using a large dataset containing a mixture of publicly available datasets. This modification enables the model to extract both global and local temporal representations tailored to animal movements. The GLTR-head can also be replaced by a frame-based majority vote head to compare the video-based approach with the image-based.

To further enhance Re-ID performance, we incorporate pose embeddings derived from a pose estimation model called SuperAnimal trained on animal datasets [6] and using DeepLabCut [7] to create a labeled subset to fine-tune the pose models. Pose-based features capture species-specific motion characteristics that remain invariant to changes in appearance caused by environmental factors such as lighting or wet coat. We aim to determine whether an animal can be re-identified based solely on its pose and whether concatenating a pose embedding with an RGB embedding improves performance. One pose recognition model from the human domain called GaitGraph learns person-specific motion patterns from skeleton sequences using spatio-temporal graph convolutional networks [8]. MotionBERT utilizes a transformer-based backbone to capture long-range dependencies in human motion, generating detailed pose representations [9]. The model employs different heads for various tasks, and we will adapt the action recognition head for re-identification. The study will evaluate the proposed models on the three previously mentioned video-based Re-ID datasets, including controlled settings and in-the-wild footage. We consider essential metrics such as rank-based retrieval accuracy, mean average precision (mAP), and robustness across various environmental conditions, including rain, darkness, different distance from the camera. The datasets include long-tailed distributions in which certain individuals appear infrequently, posing additional challenges in model generalization. Our framework addresses these challenges by applying spatial augmentations (random cropping, horizontal flipping, color jittering) and temporal augmentations (sampling random frames in order or selecting a random frame with consecutive ones) to enhance robustness in video-based re-identification. The model optimizes feature discrimination and classification using a combination of cross-entropy and triplet loss, with hyper-parameters allowing flexible weighting of both losses.

The project will involve the following key steps:

1. Literature review on video-based Re-ID, pose estimation, and domain-informed learning. We will read key papers to critically analyze recent advancements in Transformer-based Re-ID models, pose embedding extraction, and domain-specific learning techniques.

[1] https://bearresearch.org/

2. Design and implementation of a preprocessing pipeline. The pipeline will use the mentioned pre-processed datasets with a uniform dataloader. We will follow the DeepLabCut pipline that includes labeling key-points, train pose models based on the labeled poses, and evaluate their performance. The extracted pose keypoints will be integrated into existing datasets and made accessible in the DataLoader.

3. Development and comparison of multiple Re-ID models. We will compare the performance of PIT, GLTR, and GLTR with MIEW-ID against pose-enhanced architectures such as MotionBERT and Gait Graph. Additionally, we will investigate the impact of combining these models by concatenating their embeddings and measuring improvements in recognition performance.

4. Performance evaluation across datasets. The models will be evaluated on the three mentioned datasets, focusing on accuracy, robustness, and interpretability through methods such as the activation map technique [10] and embedding analysis (t-SNE plots).

The thesis will provide a detailed description of all developed methodologies, supported by extensive experimental analysis. Implementation code will be made publicly available to facilitate further research and development in the field. By integrating temporal dynamics, pose estimation, and domain knowledge, our work aims to advance automated animal Re-ID, offering valuable tools for wildlife monitoring and conservation efforts.

Advisors: Matthias Zuerl
Jonas Mueller
Prof. Dr. Bjoern Eskofier
Student: Jan Petermann
Start: 01.02.2025

References

[1] Zuerl, M.; Dirauf, R.; Koeferl, F.; Steinlein, N.; Sueskind, J.; Zanca, D.; Brehm, I.; Fersen, L.v.; Eskofier, B. PolarBearVidID: A Video-Based Re-Identification Benchmark Dataset for Polar Bears. Animals 2023, 13, 801. https://doi.org/10.3390/ani13050801

[2] Wing Print: Automated Bat Re-Identification Based On Distinct Wing Membrane Patterns (2023), Master Thesis Computer Science, Julian Deyerler Advisors: Matthias Zuerl, Philipp Schlieper, Dr. Ralph Simon, Prof. Dr. Bjoern Eskofier

[3] Zang, Xianghao and Li, Ge and Gao, Wei (2022) Multidirection and Multiscale Pyramid in Transformer for Video-Based Pedestrian Retrieval.

[4] Jianing Li, Jingdong Wang, Qi Tian, Wen Gao, Shiliang Zhang (2019) Global-Local Temporal Representations For Video Person Re-Identification

[5] Lasha Otarashvili, Tamilselvan Subramanian, Jason Holmberg, J.J. Levenson, Charles V. Stewart (2024) Multispecies Animal Re-ID Using a Large Community-Curated Dataset

[6] Shaokai Ye, Anastasiia Filippova, Jessy Lauer, Steffen Schneider, Maxime Vidal, Tian Qiu, Alexander Mathis & Mackenzie Weygandt Mathis (2024) SuperAnimal pretrained pose estimation models for behavioral analysis

[7] Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature neuroscience, 21(9), 1281-1289.

[8] Torben Teepe, Johannes Gilg, Fabian Herzog, Stefan Hörmann, Gerhard Rigoll (2022) Towards a Deeper Understanding of Skeleton-based Gait Recognition

[9] Wentao Zhu, Xiaoxuan Ma, Zhaoyang Liu, Libin Liu, Wayne Wu, Yizhou Wang (2022) MotionBERT: A Unified Perspective on Learning Human Motion Representations

[10] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. (2016) Learning deep features for discriminative localization.