Addressing the Elephant in the Room:
Robust Animal Re-Identification with Unsupervised Part-Based Feature Alignment

Code :
https://github.com/Chloe-Yu/Animal-Re-ID

Yingxue Yu, Vidit Vidit, Andrey Davydov, Martin Engilberge, Pascal Fua
École Polytechnique Fédérale de Lausanne (EPFL)

CVPR
SEATTLE, WA JUNE 17-21, 2024

Introduction

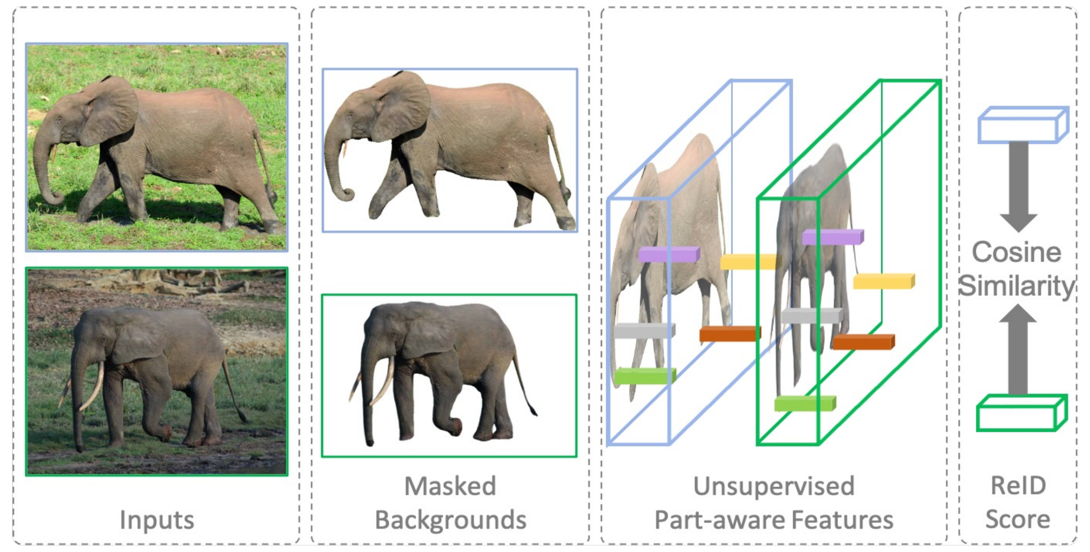
Motivation

Problem with Animal Re-ID Tasks

- Datasets scarce & lack of diversity ⇒ Background Overfitting
- Great variations within an individual
- Small variations between different individuals
- Few part-labeled images

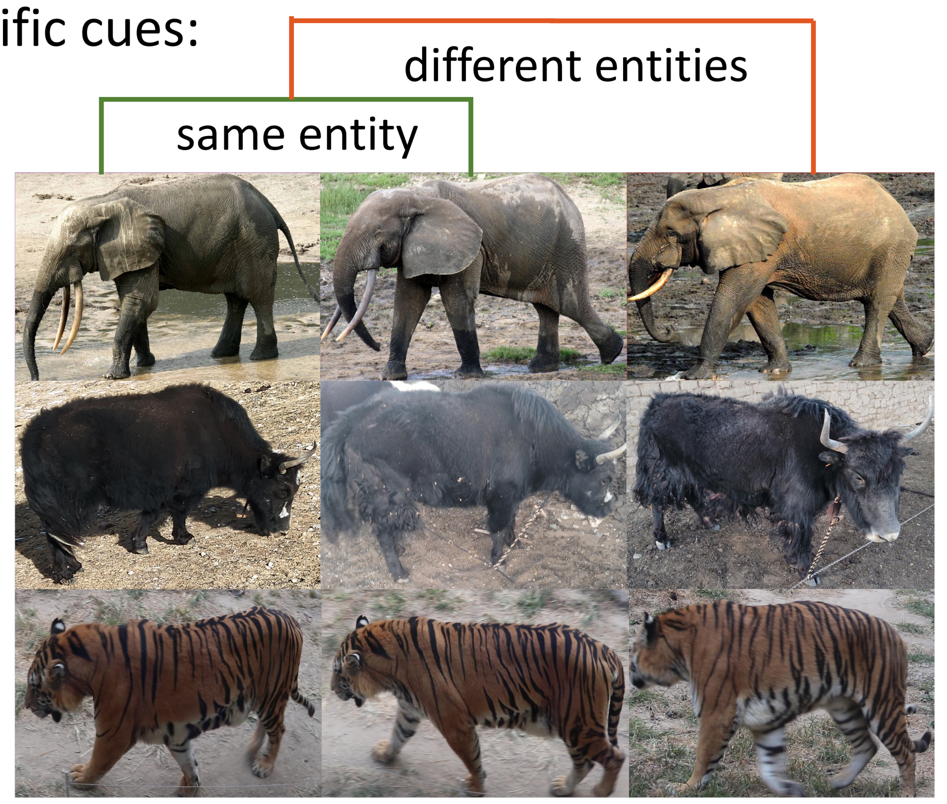
Our Solution

- Unsupervised Part-aware model on Background-Free images



Learning Part-aware Features

- Subtle species specific cues:
  - tusk
  - horn
  - stripes



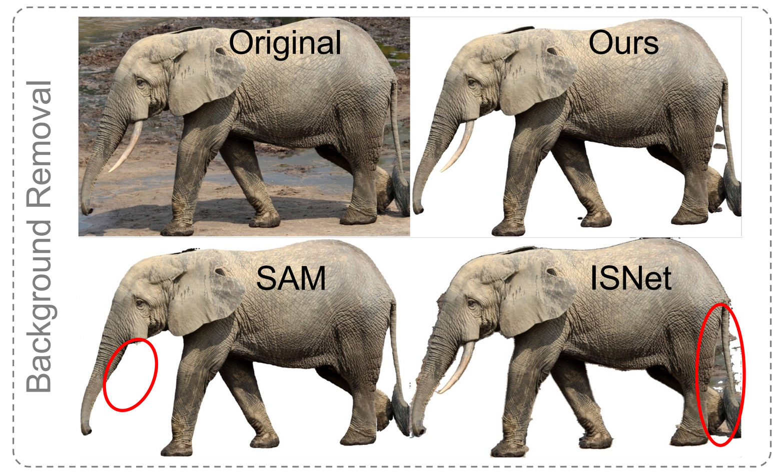
Background Removal

Background overfitting:

- Existing methods overly rely on background
- Clustering shows:
  - Grouping by background
  - Multiple IDs in the the same cluster

Our Solution

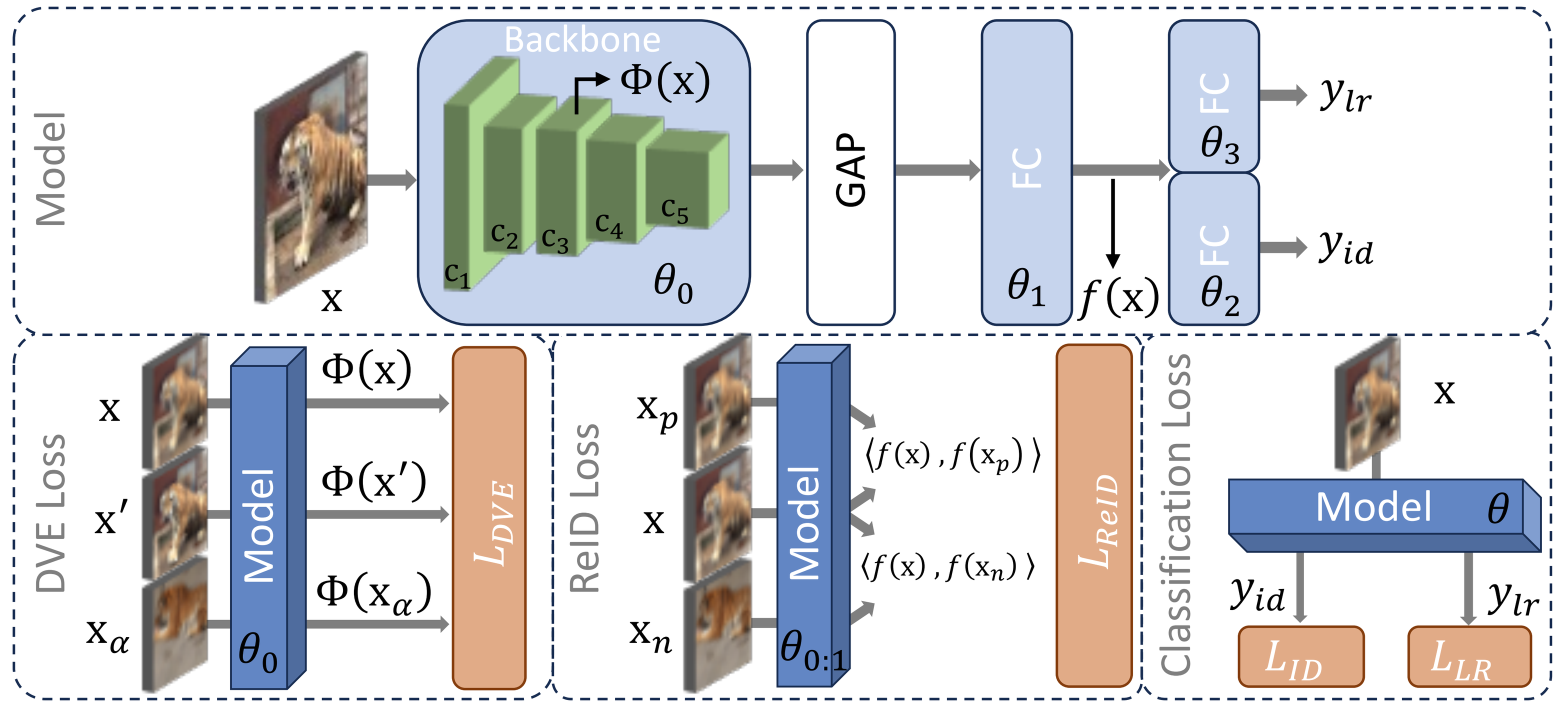
- Remove background both at training and testing time
- True evaluation of identity



Masked Backgrounds		ATRW		
Train	Test	mmAP	R@1(s)	R@1(c)
		74.5	95.7	90.3
	✓	60.2	88.6	83.4
✓	✓	66.9	90.8	86.3
✓		73.4	96.9	89.7

Model Architecture

- Joint training
- Three losses
  - Part alignment - DVE
  - ReID - Circle loss
  - Attribute classification
    - Orientation
    - ID



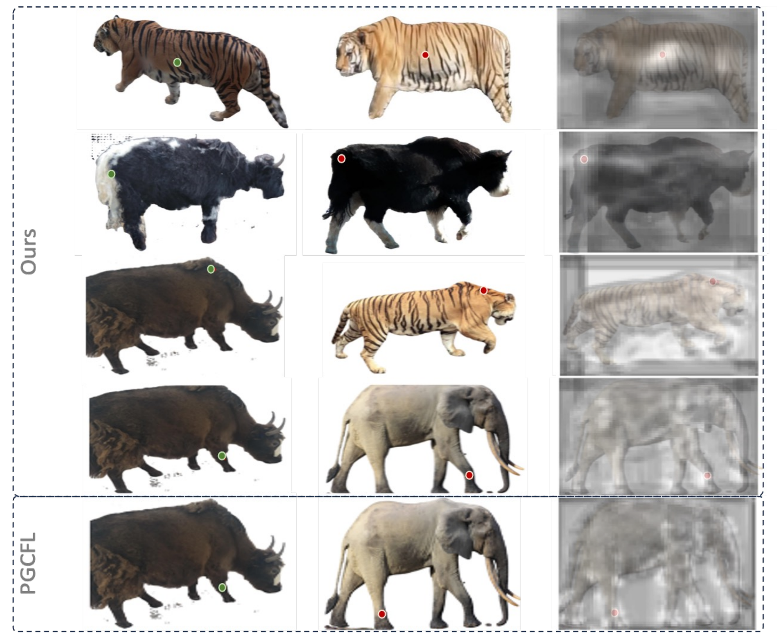
L_{DVE}: Unsupervised technique for feature alignment across body parts and pose variations [1]

Evaluation

Comparison with SOTA

Methods	Org. Re-ID Task	Pose GT	ATRW mmAP	ATRW R@1(s)	ATRW R@1(c)	YakReID-103 mAP	YakReID-103 R@1	ELPephants mAP	ELPephants R@1
CLIP-Re-ID ViT [2]	Person	✗	56.5	86.3	72.0	49.8	82.2	12.7	25.3
PPGNet R-50 [3]	Animal	✓	68.3	81.2	81.1	-	-	-	-
ResNet50 [4]	Animal	✗	65.9	91.1	83.4	60.9	86.0	20.0	33.9
ViT [5]	Animal	✗	65.5	90.3	79.4	**61.3**	88.4	20.6	36.8
PGCFL [6]	Animal	✗	66.9	90.8	**86.3**	55.8	82.7	18.5	33.4
Ours	Animal	✗	**68.6**	**92.0**	84.6	61.0	**89.4**	**24.3**	**38.7**

Feature Visualization



Ablation Study

Model Components					ATRW	
L_{DVE}	L_{ID}	L_{ReID}	L_{LR}	B.S.	mAP	R@1(c)
	✓				54.6	74.9
		✓			56.3	72.0
	✓	✓			54.4	69.7
	✓		✓		61.4	81.1
		✓	✓		60.5	76.6
	✓	✓	✓		63.4	78.9
✓	✓	✓	✓		63.4	78.3
	✓	✓	✓	✓	63.2	80.6
✓	✓	✓	✓	✓	64.9	82.9

References

[1] Thewlis, J., Albanie, S., Bilen, H., & Vedaldi, A. Unsupervised Learning of Landmarks by Descriptor Vector Exchange. *2019 ICCV*

[2] Siyuan Li, Li Sun, and Qingli Li. Clip-reid: exploiting vision-language model for image re-identification without concrete text labels. AAAI 2023

[3] Cen Liu, Rong Zhang, and Lijun Guo. Part-pose guided amur tiger re-identification. CVPR 2019.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CVPR 2016

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv

[6] Ning Liu, Qijun Zhao, Nan Zhang, Xinhua Cheng, and Jianing Zhu. Pose-guided complementary features learning for amur tiger re-identification. CVPR workshops, 2019