The 38th Annual AAAI Conference on Artificial Intelligence

UniAP: Towards Universal Animal Perception in Vision via Few-shot Learning

Meiqi Sun[1*], Zhonghan Zhao[2*], Wenhao Chai[3*], Hanjun Luo[1], Shidong Cao[1], Yanting Zhang[4], Jenq-Neng Hwang[3], Gaoang Wang[1,2,5†]

1 Zhejiang University-University of Illinois Urbana Champaign Institute, Zhejiang University

2 College of Computer Science and Technology, Zhejiang University 3 Electrical and Computer Engineering Department, University of Washington

4 Department of Computer Science and Technology, Donghua University 5 Shanghai Artificial Intelligence Laboratory

Motivation

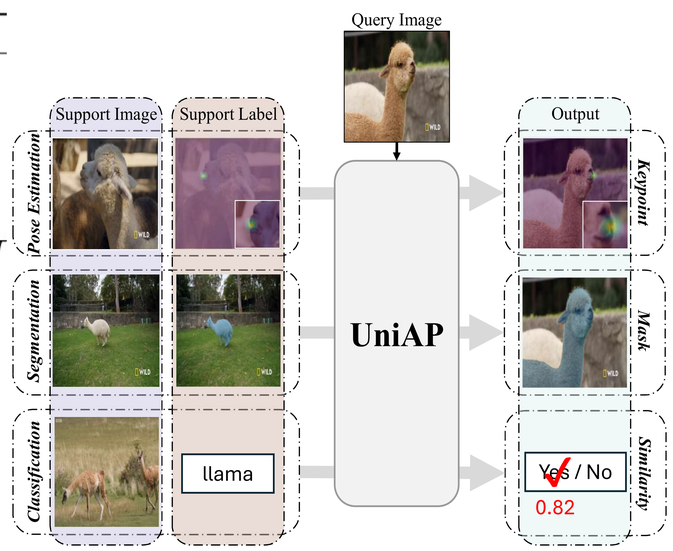
Animal visual perception is important for automatically monitoring animal health, understanding animal behaviors, and assisting animal-related research. However, it is challenging to design a perception model that can adapt to different animals across various perception tasks, due to the varying poses of a large diversity of animals, lacking data on rare species, and the semantic inconsistency of different tasks. We introduce UniAP, a novel Universal Animal Perception model that leverages few-shot learning to enable cross-species perception among various visual tasks.

Method

Algorithm 1: Universal Animal Perception (UniAP)

Input: Query Image X_q, Prompt Set $\mathcal{P}$
Output: Predictions $\hat{Y}^q$

Constants:
  $\mathcal{T} = \{\text{Pose Estimation, Semantic Segmentation}\}$
Functions:
  $f_{\mathcal{T}}$: Image encoder with task-specific parameters $\theta_{\mathcal{T}}$
  g: Label encoder (shared across tasks)
  $\mathcal{M}$: Matching module
  h: Label decoder
  σ: Similarity function
Initialize:
  $\mathbf{q} \leftarrow f_{\mathcal{T}}(X_q)$
for each (X_i^p, Y_i^p) in $\mathcal{P}$ **do**
  $\mathbf{k}_i \leftarrow f_{\mathcal{T}}(X_i^p)$ {Encode using Image Encoder}
  $\mathbf{v}_i \leftarrow g(Y_i^p)$ {Encode using Label Encoder}
end for
$\mathbf{m} \leftarrow \mathcal{M}(\mathbf{q}, \mathbf{k}_i, \mathbf{v}_i)$
$\hat{Y}^q \leftarrow h(\mathbf{m})$ {Decode using Label Decoder}
return $\hat{Y}^q$



UniAP unifies different tasks under a single model via few-shot learning.

Algorithm 2: Matching module $\mathcal{M}$ in UniAP

Input: $\mathbf{q}, \mathbf{k}, \mathbf{v}$
Output: $\mathbf{m}$

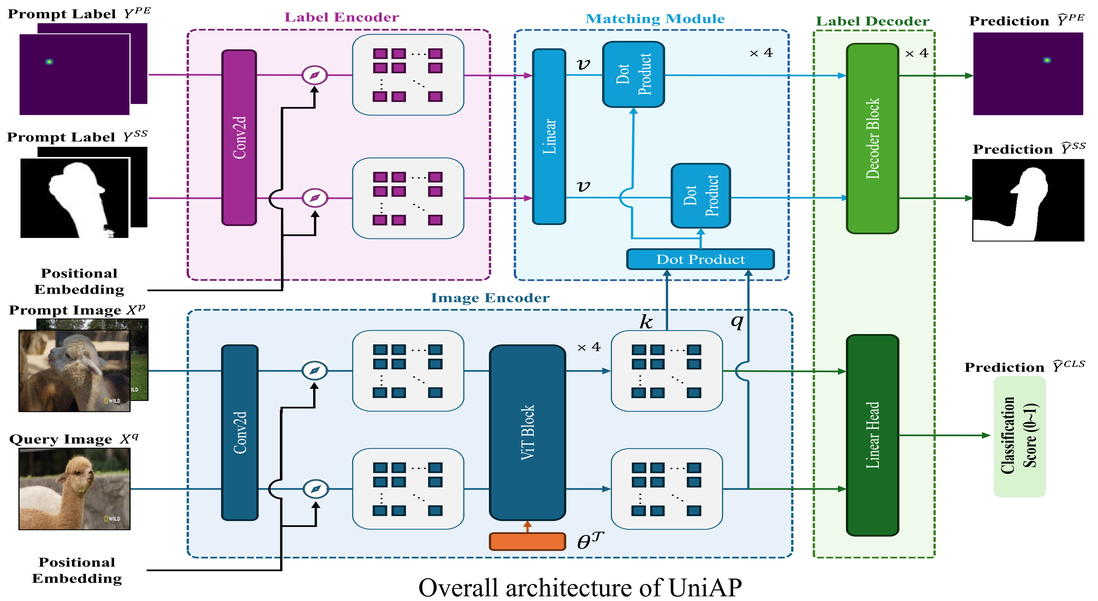
Constants:
  Number of heads H, Head size d_H
  Trainable projection matrices w_h^Q, w_h^K, w_h^V, w^O
for $h = 1$ to H **do**
  $M_A \leftarrow \frac{\mathbf{q}w_h^Q(\mathbf{k}w_h^K)^\top}{\sqrt{d_H}}$
  $\mathbf{o}_h \leftarrow \text{Softmax}(M_A)\mathbf{v}w_h^V$
end for
$\mathbf{m} \leftarrow \text{Concat}(\mathbf{o}_1, ..., \mathbf{o}_H)w^O$
return $\mathbf{m}$

Model	# tasks	# shots	Task-Specific	Shared
POMNet	one	multiple	58.25M	0
Painter	multiple	one	0	353.55M
Ours	multiple	multiple	0.07M	111.01M

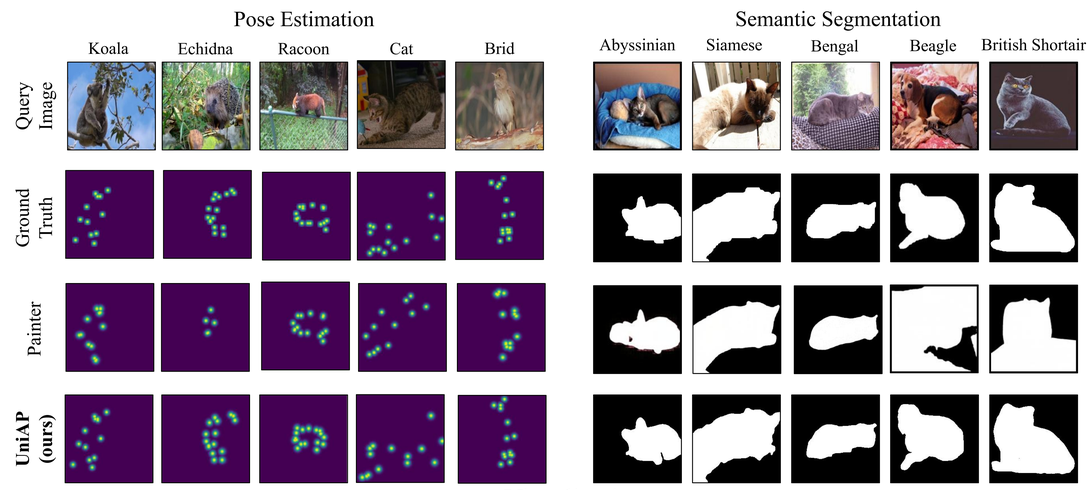
Number of task-specific and shared parameters for a task.

Dataset	bias tuning (awl+ft)	full tuning (awl+fft)
Animal Kingdom	3,119	4,369
Animal Pose	2,109	2,565
APT-36K	2,731	4,005

Number of tuning batches for bias tuning and full tunning.



Overall architecture of UniAP



Visualization

Experiment

Model	# shots	Animal Kingdom		Animal Pose		APT-36K	
		PCK@0.2	PCK@0.05	PCK@0.2	PCK@0.05	PCK@0.2	PCK@0.05
HRNet_{w48}	-	90.49	62.04	**90.47**	75.91	91.65	66.26
Painter	1	70.52	48.34	77.86	53.85	74.11	51.39
POMNet	1	59.97	30.65	73.28	51.81	63.90	38.52
	3 / 3 / 2	79.15	52.88	77.7	49.96	5.79	38.4
UniAP (ours)	1	64.44	34.73	76.67	47.31	85.31	61.72
	30 / 35 / 40	**99.65**	**98.59**	90.10	**77.78**	**96.47**	**86.18**

Results of Pose Estimation

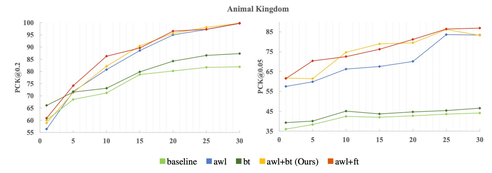
Model	# shots	Acc.	mIoU
SAM_{user}	-	92.06	88.99
Painter	1	86.47	77.72
UniAP (ours)	1	97.08	93.38
	10	**97.11**	**94.27**

Model	# shots	Acc.
$\text{CLIP}_{\text{ViT-Base}}$	-	88.54
$\text{CLIP}_{\text{ViT-Large}}$	-	90.37
UniAP (ours)	1	92.19
	5	**93.13**

Results of Semantic Segmentation and classification

Setting	# shots	Animal Kingdom		Animal Pose		APT-36K	
		PCK@0.2	PCK@0.05	PCK@0.2	PCK@0.05	PCK@0.2	PCK@0.05
OOD	1	64.44	34.73	76.67	47.31	85.31	61.72
	30 / 35 / 40	**99.65**	**98.59**	90.10	77.78	96.47	86.18
ID	1	77.39	61.25	77.69	46.74	92.41	61.41
	20 / 20 / 35	99.26	98.10	**94.97**	**88.47**	**96.92**	**92.70**
CE	1	54.18	24.61	60.47	28.47	78.39	47.92
	5 / 5 / 5	65.50	28.86	77.33	50.62	88.72	71.67

Ablation studies on evaluation settings.



Ablation studies on the performance of various shots