Supplementary Material

LiDeRe: A Lightweight Readout for Fast and Data-Efficient Dense Prediction

Timo Lüddecke[1], Jan Frederik Meier[1], Jan van Delden[1], Alexander Ecker[1,2]

[1]University of Göttingen [2]MPI for Dynamics and Self-Organization

https://eckerlab.org/code/lidere

Pose Estimation Implementation Details

For the pose experiments, we use the iRodent and Horse-30 parts of SuperAnimal Quadruped dataset and the provided training-test splits. The dataset is available here: https://zenodo.org/records/14016777

Following the protocol of SuperAnimal, we evaluate on cropped animals (bounding boxes are provided in the dataset). RMSE scores are computed in the original image space. For mAP computation we use pycocotools and set detections scores to 1 for all detections.

Hyperparameters

Below we provide the hyperparameters for all datasets.

	n_iterations	bs	lr	val_interval	wd	amp	cos_lr
Leaf	500	32	0.001	50	0.01	✓	✓
Camo	1000	32	0.001	100	0.01	✓	-
Trans10k	3000	32	0.001	250	0.01	✓	-
iRodent	2500	16	0.001	-	0.01	✓	✓
Horse10	10000	16	0.001	-	0.01	✓	✓
PlantDoc	7000	32	0.001	250	0.01	✓	-
BSDS500	3000	32	0.001	250	0.01	✓	✓

Table 1: Hyperparameters of the experiments.

Model Details: Implicit MLP

Input (4D) for position $\mathbf{p}^F \in \mathbb{R}^2$ in the feature volume and query position $\mathbf{p}^T \in \mathbb{R}^2$, which are both normalized to be in $[0, 1]$, the MLP receives the following input vector $[\mathbf{p}_0^F - \mathbf{p}_0^T, \mathbf{p}_1^F - \mathbf{p}_1^T, (\mathbf{p}_0^F - \mathbf{p}_0^T)^2, (\mathbf{p}_1^F - \mathbf{p}_1^T)^2]$.
The implicit MLP has this architecture:

```
[Linear(4, 32), Sine, Linear(32, 32), Sine, Linear(32, 8)]
```

Visualizations

We visualize the prediction of our model with different backbones and the linear evaluation baseline (Figure 1).

Repetitions

To increase the reliability of our reported scores, we average the scores of three training runs for all detection (Table 2) and pose estimation experiments (Tables 3). Below all scores are reported.

	mAP ↑			mAP50 ↑		
	1	2	3	1	2	3
ours	49.6	45	46.9	70.9	66.3	68.9
ours + LoRA	48.2	45.3	47.5	68.6	68.0	66.8
ours (ViT-L)	49.7	52.1	51	72.6	75.5	72.6
ours (ViT-L) + LoRA	48.2	53.5	51.5	66.2	73.1	69.1
no interp. prior (**I**)	40.5	40.8	42.4	62.9	64.8	65.4
no feedforward (FF)	45.4	46.3	43.9	65.3	66.4	65.5
no content-guided att. (A_C)	42.7	41.8	44.1	62.5	63.1	64.9
none (I, FF, A_C)	39.0	38.1	34.2	61.8	60.8	57.5
Lin. eval	24.9	25.1	26.2	50.6	51.9	51.1
Lin. eval + LoRA	26.3	26	25.7	52.9	51.5	51.8

Table 2: Three repetitions for object detection on PlantDoc

	iRodent						Horse10					
	RMSE ↓			mAP ↑			RMSE ↓			mAP ↑		
	1	2	3	1	2	3	1	2	3	1	2	3
ours	30.7	30.0	29.9	70.4	71.0	70.6	3.12	3.43	3.24	87.9	86.2	87.0
ours + LoRA	30.5	30.0	28.9	71.6	73.6	73.3	1.78	1.73	1.71	93.8	94.5	94.8
ours (ViT-L)	27.8	28.2	29.4	73.9	72.9	71.6	2.89	2.21	2.17	90.4	93.4	93.4
ours (ViT-L) + LoRA	28.0	27.1	28.8	72.7	73.7	75.0	1.58	1.55	1.57	96.5	96.4	96.0
no interp. prior (**I**)	31.9	34.2	34.0	68.0	67.9	67.4	4.61	4.5	4.13	81.6	82.1	83.5
no feedforward (FF)	33.5	32.0	31.7	65.2	68.8	67.5	3.33	2.95	3.06	87.0	87.5	87.3
no content-guided att. (A_C)	30.3	30.5	31.4	70.2	71.0	69.3	3.35	3.56	3.67	86.9	85.6	85.4
none (I, FF, A_C)	35.6	34.3	34.5	61.4	62.1	63.5	4.37	4.12	4.42	81.9	82.3	82.2
Lin. eval	52.4	52.4	52.7	47.0	47.3	46.5	7.41	7.41	7.4	68.0	67.8	67.9
Lin. eval + LoRA	39.1	35.5	33.2	56.2	61.9	64.1	2.31	2.43	2.37	89.4	88.7	89.7

Table 3: Three repetitions for pose estimation.

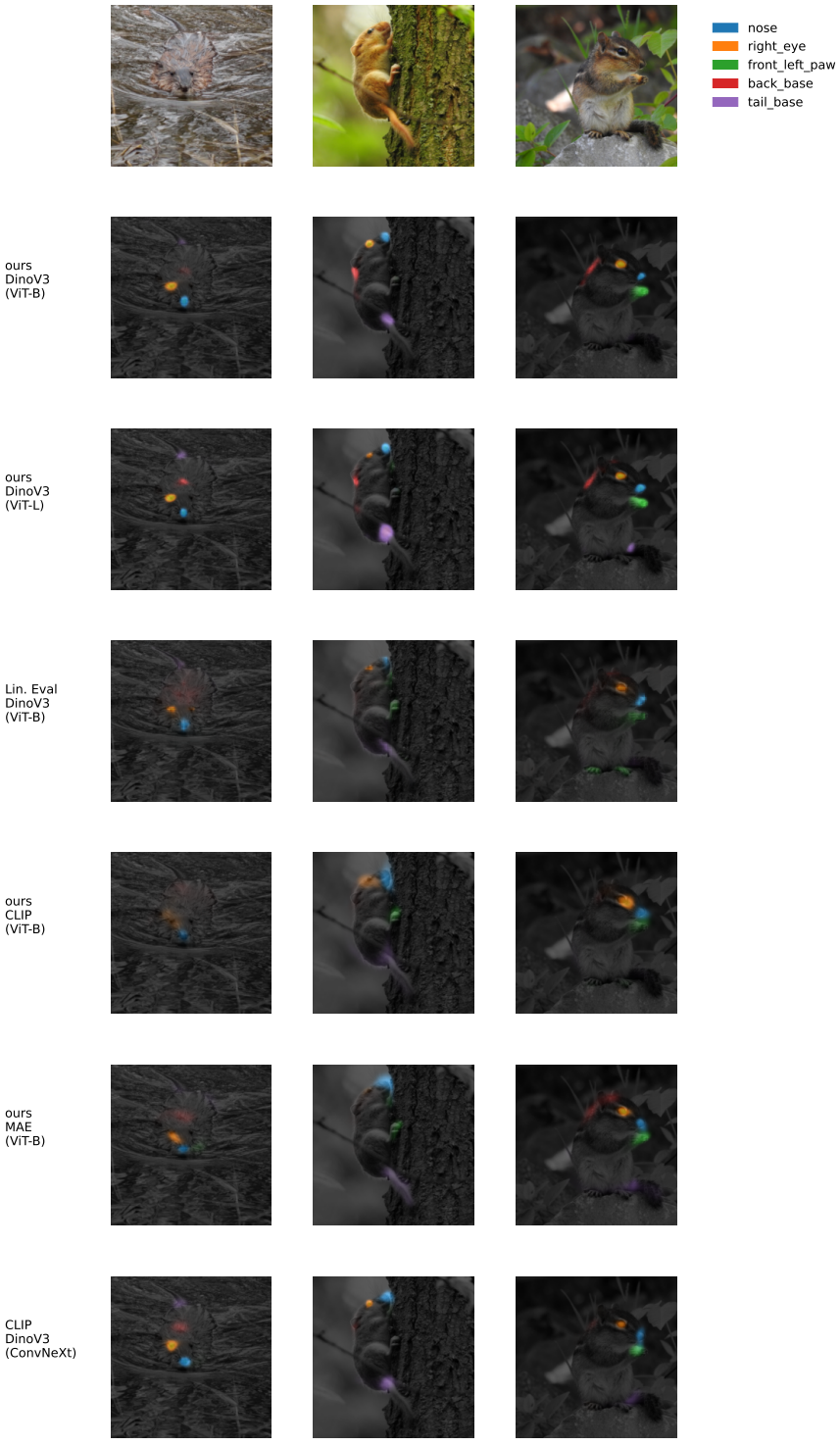

Figure 1: Visualization of the keypoint heatmaps of the iRodent dataset. Each color corresponds to a different keypoint.