AP-10K: A Benchmark for Animal Pose Estimation in the Wild: Supplementary Material

Hang Yu[1*], Yufei Xu[2*], Jing Zhang[2†], Wei Zhao[1†], Ziyu Guan[1], Dacheng Tao[2,3]

[1]Xidian University, China, [2]The University of Sydney, Australia, [3]JD Explore Academy, China

A Appendix

A.1 More quantitative and qualitative results

Table S1: The complete results of different models on the validation set of the SL track.

Pretraining on ImageNet	AP	$\mathrm{AP}_{.5}$	$\mathrm{AP}_{.75}$	AP_{M}	AP_{L}
HRNet-w32 [13]	$0.738_{\pm 0.006}$	$0.958_{\pm 0.004}$	$0.808_{\pm 0.007}$	$0.592_{\pm 0.074}$	$0.743_{\pm 0.006}$
HRNet-w48 [13]	$0.744_{\pm 0.004}$	$0.959_{\pm 0.002}$	$0.807_{\pm 0.006}$	$0.589_{\pm 0.042}$	$0.748_{\pm 0.003}$
ResNet50 [7]	$0.699_{\pm 0.004}$	$0.940_{\pm 0.003}$	$0.760_{\pm 0.003}$	$0.570_{\pm 0.078}$	$0.703_{\pm 0.004}$
ResNet101 [7]	$0.698_{\pm 0.002}$	$0.943_{\pm 0.002}$	$0.754_{\pm 0.007}$	$0.543_{\pm 0.070}$	$0.702_{\pm 0.001}$
Hourglass [10]	$0.729_{\pm 0.001}$	$0.951_{\pm 0.002}$	$0.793_{\pm 0.002}$	$0.606_{\pm 0.042}$	$0.733_{\pm 0.001}$
Training from scratch	AP	$\mathrm{AP}_{.5}$	$\mathrm{AP}_{.75}$	AP_{M}	AP_{L}
HRNet-w32 [13]	$0.703_{\pm 0.002}$	$0.940_{\pm 0.001}$	$0.764_{\pm 0.004}$	$0.599_{\pm 0.020}$	$0.707_{\pm 0.003}$
HRNet-w48 [13]	$0.711_{\pm 0.003}$	$0.943_{\pm 0.002}$	$0.771_{\pm 0.004}$	$0.604_{\pm 0.047}$	$0.715_{\pm 0.003}$
ResNet50 [7]	$0.646_{\pm 0.001}$	$0.913_{\pm 0.006}$	$0.694_{\pm 0.007}$	$0.522_{\pm 0.053}$	$0.649_{\pm 0.002}$
ResNet101 [7]	$0.667_{\pm 0.002}$	$0.924_{\pm 0.003}$	$0.715_{\pm 0.003}$	$0.542_{\pm 0.042}$	$0.671_{\pm 0.002}$
Hourglass [10]	$0.686_{\pm 0.006}$	$0.933_{\pm 0.003}$	$0.739_{\pm 0.011}$	$0.591_{\pm 0.027}$	$0.690_{\pm 0.004}$

In this Appendix, we provide the full evaluation results of HRNet-W32 [13], HRNet-w48 [13], SimpleBaseline [15] with ResNet50 [7] and ResNet101 [7] and Hourglass [10] in terms of different evaluation metrics on the validation set of the SL Track, as shown in Table S1. It shows that the advanced network structure like HRNet [13] outperforms the other backbones on most categories in terms of all evaluation metrics. It also demonstrates that pretraining on ImageNet can accelerate the convergence speed and achieve a better performance than training from scratch.

Besides, we also present the per species results of HRNet-W32 [13] on the test set the SL Track at the settings of "Pretraining on ImageNet" and "Training from scratch" in Table S2 and Table S3, respectively. We also provide another test result which contains 56 animals for reference. *i.e.*, Dividing Dog into Chihuahua, Collie, Dog, dalmatian and German Shepherd, dividing Cat into cat and Perisian Cat and Siamese Cat. As we have mentioned in **??**, Dog and Cat varies widely in appearance owning to human's cultivation for family pets.

Comparing Table S2 with the last rows in Table **??**, Table **??** and Table **??** in the paper, where all the scores represent test performance on seen species, we can find that using the same amount of training data but from more diverse species can help the model learning better feature representation and achieve better performance, *e.g.*, the mAP for Sheep is 0.663 v.s. 0.761, which is obtained by the model trained only using the Bovidae family). This observation further confirms the value of our AP-10K dataset.

*Equal contribution. †Corresponding author.

Submitted to the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. Do not distribute.

Table S2: Per-species results of HRNet-w32 [13] on the test set the SL Track at the setting of pretraining on ImageNet.

species	AP	AP.5	AP.75	AP(M)	AP(L)
Antelope	$0.856_{\pm 0.014}$	$0.969_{\pm 0.008}$	$0.924_{\pm 0.007}$	$0.732_{\pm 0.024}$	$0.869_{\pm 0.012}$
Argali Sheep	$0.871_{\pm 0.069}$	$0.966_{\pm 0.048}$	$0.933_{\pm 0.095}$	N/A	$0.914_{\pm 0.011}$
Bison	$0.723_{\pm 0.021}$	$0.941_{\pm 0.028}$	$0.804_{\pm 0.017}$	$0.487_{\pm 0.070}$	$0.788_{\pm 0.015}$
Buffalo	$0.834_{\pm 0.045}$	$0.978_{\pm 0.032}$	$0.895_{\pm 0.040}$	$0.141_{\pm 0.813}$	$0.844_{\pm 0.039}$
Cow	$0.757_{\pm 0.019}$	$0.979_{\pm 0.015}$	$0.824_{\pm 0.042}$	$0.551_{\pm 0.035}$	$0.765_{\pm 0.020}$
Sheep	$0.761_{\pm 0.011}$	$0.959_{\pm 0.015}$	$0.833_{\pm 0.010}$	$0.598_{\pm 0.154}$	$0.769_{\pm 0.008}$
Dog	$0.771_{\pm 0.013}$	$0.962_{\pm 0.006}$	$0.836_{\pm 0.003}$	$0.129_{\pm 0.803}$	$0.772_{\pm 0.011}$
Fox	$0.780_{\pm 0.002}$	$0.943_{\pm 0.025}$	$0.843_{\pm 0.014}$	N/A	$0.780_{\pm 0.002}$
Wolf	$0.778_{\pm 0.033}$	$0.972_{\pm 0.020}$	$0.864_{\pm 0.061}$	N/A	$0.779_{\pm 0.034}$
Beaver	$0.562_{\pm 0.004}$	$0.919_{\pm 0.021}$	$0.663_{\pm 0.015}$	$0.867_{\pm 0.047}$	$0.557_{\pm 0.008}$
Alouatta	$0.784_{\pm 0.038}$	$1.000_{\pm 0.000}$	$0.924_{\pm 0.061}$	N/A	$0.784_{\pm 0.038}$
Monkey	$0.690_{\pm 0.027}$	$0.961_{\pm 0.013}$	$0.769_{\pm 0.060}$	N/A	$0.690_{\pm 0.027}$
Noisy Night Monkey	$0.765_{\pm 0.026}$	$0.960_{\pm 0.000}$	$0.803_{\pm 0.068}$	N/A	$0.777_{\pm 0.040}$
Spider Monkey	$0.602_{\pm 0.046}$	$0.871_{\pm 0.024}$	$0.627_{\pm 0.072}$	$0.300_{\pm 0.920}$	$0.600_{\pm 0.048}$
Uakari	$0.829_{\pm 0.007}$	$1.000_{\pm 0.000}$	$0.976_{\pm 0.034}$	N/A	$0.829_{\pm 0.007}$
Deer	$0.794_{\pm 0.029}$	$0.950_{\pm 0.014}$	$0.847_{\pm 0.029}$	$0.157_{\pm 0.829}$	$0.811_{\pm 0.006}$
Moose	$0.776_{\pm 0.014}$	$0.973_{\pm 0.005}$	$0.843_{\pm 0.054}$	$0.838_{\pm 0.044}$	$0.773_{\pm 0.017}$
Hamster	$0.571_{\pm 0.064}$	$0.864_{\pm 0.042}$	$0.563_{\pm 0.079}$	$0.241_{\pm 0.175}$	$0.603_{\pm 0.051}$
Elephant	$0.704_{\pm 0.040}$	$0.918_{\pm 0.010}$	$0.774_{\pm 0.051}$	N/A	$0.703_{\pm 0.041}$
Horse	$0.749_{\pm 0.028}$	$0.934_{\pm 0.038}$	$0.798_{\pm 0.036}$	N/A	$0.764_{\pm 0.022}$
Zebra	$0.755_{\pm 0.026}$	$0.921_{\pm 0.013}$	$0.775_{\pm 0.041}$	N/A	$0.764_{\pm 0.019}$
Bobcat	$0.751_{\pm 0.055}$	$0.976_{\pm 0.034}$	$0.842_{\pm 0.057}$	N/A	$0.753_{\pm 0.057}$
Cat	$0.672_{\pm 0.036}$	$0.935_{\pm 0.006}$	$0.750_{\pm 0.067}$	N/A	$0.675_{\pm 0.037}$
Cheetah	$0.735_{\pm 0.029}$	$0.936_{\pm 0.010}$	$0.800_{\pm 0.026}$	N/A	$0.769_{\pm 0.032}$
Jaguar	$0.785_{\pm 0.035}$	$0.972_{\pm 0.020}$	$0.864_{\pm 0.082}$	$0.592_{\pm 0.170}$	$0.803_{\pm 0.024}$
King Cheetah	$0.848_{\pm 0.122}$	$0.944_{\pm 0.079}$	$0.944_{\pm 0.079}$	N/A	$0.848_{\pm 0.122}$
Leopard	$0.731_{\pm 0.009}$	$0.956_{\pm 0.034}$	$0.796_{\pm 0.047}$	N/A	$0.731_{\pm 0.009}$
Lion	$0.726_{\pm 0.017}$	$0.927_{\pm 0.024}$	$0.817_{\pm 0.038}$	N/A	$0.726_{\pm 0.017}$
Panther	$0.772_{\pm 0.048}$	$0.985_{\pm 0.021}$	$0.851_{\pm 0.027}$	$0.084_{\pm 0.769}$	$0.780_{\pm 0.051}$
Snow Leopard	$0.858_{\pm 0.042}$	$1.000_{\pm 0.000}$	$0.939_{\pm 0.048}$	$0.183_{\pm 0.849}$	$0.865_{\pm 0.035}$
Tiger	$0.811_{\pm 0.012}$	$0.973_{\pm 0.021}$	$0.872_{\pm 0.009}$	N/A	$0.812_{\pm 0.013}$
Giraffe	$0.803_{\pm 0.016}$	$0.958_{\pm 0.007}$	$0.842_{\pm 0.046}$	N/A	$0.808_{\pm 0.016}$
Hippo	$0.457_{\pm 0.062}$	$0.746_{\pm 0.089}$	$0.451_{\pm 0.064}$	$0.627_{\pm 0.273}$	$0.462_{\pm 0.052}$
Chimpanzee	$0.642_{\pm 0.009}$	$0.912_{\pm 0.034}$	$0.689_{\pm 0.044}$	$0.567_{\pm 0.126}$	$0.645_{\pm 0.018}$
Gorilla	$0.753_{\pm 0.010}$	$0.950_{\pm 0.029}$	$0.881_{\pm 0.038}$	N/A	$0.753_{\pm 0.010}$
Rabbit	$0.731_{\pm 0.013}$	$0.968_{\pm 0.003}$	$0.821_{\pm 0.023}$	N/A	$0.731_{\pm 0.013}$
Skunk	$0.558_{\pm 0.016}$	$0.864_{\pm 0.035}$	$0.652_{\pm 0.020}$	N/A	$0.562_{\pm 0.018}$
Mouse	$0.640_{\pm 0.015}$	$0.954_{\pm 0.033}$	$0.717_{\pm 0.036}$	$0.123_{\pm 0.796}$	$0.640_{\pm 0.013}$
Rat	$0.650_{\pm 0.031}$	$0.929_{\pm 0.015}$	$0.729_{\pm 0.052}$	$0.200_{\pm 0.849}$	$0.652_{\pm 0.030}$
Otter	$0.559_{\pm 0.017}$	$0.909_{\pm 0.031}$	$0.582_{\pm 0.050}$	N/A	$0.559_{\pm 0.017}$
Weasel	$0.716_{\pm 0.025}$	$0.955_{\pm 0.018}$	$0.790_{\pm 0.016}$	N/A	$0.717_{\pm 0.025}$
Raccoon	$0.629_{\pm 0.032}$	$0.954_{\pm 0.005}$	$0.675_{\pm 0.061}$	$0.401_{\pm 0.143}$	$0.651_{\pm 0.023}$
Rhino	$0.819_{\pm 0.021}$	$0.980_{\pm 0.000}$	$0.892_{\pm 0.048}$	N/A	$0.819_{\pm 0.021}$
Marmot	$0.784_{\pm 0.032}$	$1.000_{\pm 0.000}$	$0.860_{\pm 0.053}$	N/A	$0.784_{\pm 0.032}$
Squirrel	$0.845_{\pm 0.008}$	$1.000_{\pm 0.000}$	$0.943_{\pm 0.005}$	N/A	$0.845_{\pm 0.008}$
Pig	$0.642_{\pm 0.033}$	$0.964_{\pm 0.026}$	$0.677_{\pm 0.066}$	N/A	$0.640_{\pm 0.033}$
Black Bear	$0.662_{\pm 0.040}$	$0.967_{\pm 0.047}$	$0.865_{\pm 0.128}$	N/A	$0.662_{\pm 0.040}$
Brown Bear	$0.617_{\pm 0.029}$	$0.940_{\pm 0.011}$	$0.660_{\pm 0.051}$	$0.067_{\pm 0.759}$	$0.620_{\pm 0.026}$
Panda	$0.583_{\pm 0.008}$	$0.912_{\pm 0.020}$	$0.607_{\pm 0.082}$	N/A	$0.583_{\pm 0.009}$
Polar Bear	$0.631_{\pm 0.016}$	$0.948_{\pm 0.054}$	$0.659_{\pm 0.056}$	N/A	$0.631_{\pm 0.016}$

Table S3: Per-species results of HRNet-w32 [13] on the test set the SL Track at the setting of training from scratch.

species	AP	AP.5	AP.75	AP(M)	AP(L)
Antelope	$0.835_{\pm 0.019}$	$0.955_{\pm 0.021}$	$0.911_{\pm 0.021}$	$0.706_{\pm 0.026}$	$0.851_{\pm 0.020}$
Argali Sheep	$0.858_{\pm 0.065}$	$0.966_{\pm 0.048}$	$0.932_{\pm 0.097}$	N/A	$0.904_{\pm 0.005}$
Bison	$0.705_{\pm 0.026}$	$0.917_{\pm 0.036}$	$0.773_{\pm 0.020}$	$0.484_{\pm 0.115}$	$0.769_{\pm 0.014}$
Buffalo	$0.810_{\pm 0.044}$	$0.973_{\pm 0.039}$	$0.870_{\pm 0.060}$	$0.136_{\pm 0.806}$	$0.818_{\pm 0.043}$
Cow	$0.733_{\pm 0.018}$	$0.962_{\pm 0.017}$	$0.810_{\pm 0.009}$	$0.571_{\pm 0.091}$	$0.742_{\pm 0.018}$
Sheep	$0.741_{\pm 0.011}$	$0.950_{\pm 0.020}$	$0.797_{\pm 0.013}$	$0.593_{\pm 0.122}$	$0.748_{\pm 0.014}$
Dog	$0.743_{\pm 0.016}$	$0.960_{\pm 0.014}$	$0.807_{\pm 0.011}$	$0.100_{\pm 0.787}$	$0.745_{\pm 0.013}$
Fox	$0.753_{\pm 0.028}$	$0.917_{\pm 0.022}$	$0.792_{\pm 0.036}$	N/A	$0.753_{\pm 0.028}$
Wolf	$0.756_{\pm 0.030}$	$0.966_{\pm 0.025}$	$0.771_{\pm 0.053}$	N/A	$0.756_{\pm 0.030}$
Beaver	$0.483_{\pm 0.026}$	$0.848_{\pm 0.024}$	$0.465_{\pm 0.111}$	$0.384_{\pm 0.327}$	$0.490_{\pm 0.029}$
Alouatta	$0.745_{\pm 0.074}$	$0.970_{\pm 0.042}$	$0.896_{\pm 0.097}$	N/A	$0.745_{\pm 0.074}$
Monkey	$0.621_{\pm 0.025}$	$0.906_{\pm 0.030}$	$0.655_{\pm 0.056}$	N/A	$0.623_{\pm 0.027}$
Noisy Night Monkey	$0.722_{\pm 0.031}$	$0.935_{\pm 0.034}$	$0.764_{\pm 0.046}$	N/A	$0.721_{\pm 0.030}$
Spider Monkey	$0.514_{\pm 0.043}$	$0.824_{\pm 0.062}$	$0.513_{\pm 0.054}$	$0.300_{\pm 0.920}$	$0.513_{\pm 0.045}$
Uakari	$0.750_{\pm 0.004}$	$1.000_{\pm 0.000}$	$0.886_{\pm 0.029}$	N/A	$0.750_{\pm 0.004}$
Deer	$0.805_{\pm 0.032}$	$0.967_{\pm 0.034}$	$0.856_{\pm 0.022}$	$0.134_{\pm 0.806}$	$0.822_{\pm 0.011}$
Moose	$0.741_{\pm 0.007}$	$0.939_{\pm 0.03}$	$0.815_{\pm 0.011}$	$0.776_{\pm 0.055}$	$0.739_{\pm 0.011}$
Hamster	$0.543_{\pm 0.031}$	$0.833_{\pm 0.054}$	$0.553_{\pm 0.017}$	$0.209_{\pm 0.086}$	$0.574_{\pm 0.019}$
Elephant	$0.664_{\pm 0.035}$	$0.881_{\pm 0.051}$	$0.725_{\pm 0.064}$	N/A	$0.662_{\pm 0.037}$
Horse	$0.759_{\pm 0.033}$	$0.928_{\pm 0.043}$	$0.802_{\pm 0.025}$	$0.023_{\pm 0.731}$	$0.765_{\pm 0.033}$
Zebra	$0.727_{\pm 0.025}$	$0.886_{\pm 0.014}$	$0.741_{\pm 0.027}$	N/A	$0.736_{\pm 0.018}$
Bobcat	$0.746_{\pm 0.044}$	$0.983_{\pm 0.024}$	$0.812_{\pm 0.058}$	N/A	$0.746_{\pm 0.044}$
Cat	$0.638_{\pm 0.030}$	$0.926_{\pm 0.023}$	$0.693_{\pm 0.053}$	N/A	$0.639_{\pm 0.031}$
Cheetah	$0.707_{\pm 0.043}$	$0.923_{\pm 0.018}$	$0.762_{\pm 0.074}$	N/A	$0.739_{\pm 0.038}$
Jaguar	$0.764_{\pm 0.038}$	$0.959_{\pm 0.032}$	$0.834_{\pm 0.039}$	$0.662_{\pm 0.225}$	$0.775_{\pm 0.027}$
King Cheetah	$0.813_{\pm 0.074}$	$0.944_{\pm 0.079}$	$0.944_{\pm 0.079}$	N/A	$0.813_{\pm 0.074}$
Leopard	$0.722_{\pm 0.026}$	$0.932_{\pm 0.020}$	$0.787_{\pm 0.024}$	N/A	$0.722_{\pm 0.026}$
Lion	$0.689_{\pm 0.012}$	$0.916_{\pm 0.018}$	$0.738_{\pm 0.017}$	N/A	$0.689_{\pm 0.012}$
Panther	$0.760_{\pm 0.024}$	$0.984_{\pm 0.023}$	$0.815_{\pm 0.024}$	$0.167_{\pm 0.829}$	$0.764_{\pm 0.025}$
Snow Leopard	$0.839_{\pm 0.036}$	$0.980_{\pm 0.028}$	$0.889_{\pm 0.005}$	$0.183_{\pm 0.849}$	$0.844_{\pm 0.030}$
Tiger	$0.764_{\pm 0.018}$	$0.971_{\pm 0.024}$	$0.828_{\pm 0.026}$	N/A	$0.764_{\pm 0.018}$
Giraffe	$0.784_{\pm 0.009}$	$0.942_{\pm 0.020}$	$0.819_{\pm 0.016}$	N/A	$0.789_{\pm 0.013}$
Hippo	$0.402_{\pm 0.026}$	$0.685_{\pm 0.030}$	$0.410_{\pm 0.016}$	$0.705_{\pm 0.206}$	$0.396_{\pm 0.023}$
Chimpanzee	$0.619_{\pm 0.031}$	$0.900_{\pm 0.049}$	$0.707_{\pm 0.026}$	$0.537_{\pm 0.140}$	$0.619_{\pm 0.042}$
Gorilla	$0.736_{\pm 0.026}$	$0.943_{\pm 0.038}$	$0.875_{\pm 0.077}$	N/A	$0.736_{\pm 0.026}$
Rabbit	$0.711_{\pm 0.045}$	$0.960_{\pm 0.008}$	$0.799_{\pm 0.072}$	N/A	$0.711_{\pm 0.045}$
Skunk	$0.499_{\pm 0.039}$	$0.830_{\pm 0.079}$	$0.514_{\pm 0.054}$	N/A	$0.504_{\pm 0.041}$
Mouse	$0.603_{\pm 0.035}$	$0.904_{\pm 0.011}$	$0.675_{\pm 0.032}$	$0.085_{\pm 0.768}$	$0.608_{\pm 0.036}$
Rat	$0.580_{\pm 0.043}$	$0.936_{\pm 0.048}$	$0.643_{\pm 0.059}$	$0.100_{\pm 0.787}$	$0.581_{\pm 0.046}$
Otter	$0.537_{\pm 0.017}$	$0.868_{\pm 0.021}$	$0.538_{\pm 0.018}$	N/A	$0.538_{\pm 0.017}$
Weasel	$0.649_{\pm 0.041}$	$0.945_{\pm 0.026}$	$0.691_{\pm 0.075}$	N/A	$0.649_{\pm 0.041}$
Raccoon	$0.585_{\pm 0.028}$	$0.940_{\pm 0.009}$	$0.590_{\pm 0.113}$	$0.363_{\pm 0.124}$	$0.601_{\pm 0.026}$
Rhino	$0.762_{\pm 0.048}$	$0.950_{\pm 0.021}$	$0.824_{\pm 0.067}$	N/A	$0.762_{\pm 0.048}$
Marmot	$0.768_{\pm 0.043}$	$1.000_{\pm 0.000}$	$0.843_{\pm 0.099}$	N/A	$0.768_{\pm 0.043}$
Squirrel	$0.849_{\pm 0.022}$	$1.000_{\pm 0.000}$	$0.962_{\pm 0.027}$	N/A	$0.849_{\pm 0.022}$
Pig	$0.612_{\pm 0.045}$	$0.907_{\pm 0.049}$	$0.648_{\pm 0.057}$	N/A	$0.610_{\pm 0.045}$
Black Bear	$0.684_{\pm 0.034}$	$0.967_{\pm 0.047}$	$0.861_{\pm 0.133}$	N/A	$0.684_{\pm 0.034}$
Brown Bear	$0.584_{\pm 0.025}$	$0.960_{\pm 0.014}$	$0.608_{\pm 0.067}$	$0.167_{\pm 0.834}$	$0.583_{\pm 0.023}$
Panda	$0.514_{\pm 0.056}$	$0.908_{\pm 0.034}$	$0.557_{\pm 0.099}$	N/A	$0.513_{\pm 0.056}$
Polar Bear	$0.619_{\pm 0.016}$	$0.928_{\pm 0.028}$	$0.650_{\pm 0.021}$	N/A	$0.619_{\pm 0.016}$

Table S4: Per-species results of HRNet-w32 [13] on the **56 animals** test set the SL Track at the setting of pretraining on ImageNet.

species	AP	$AP_{.5}$	$AP_{.75}$	AP_M	AP_L
Antelope	$0.856_{\pm 0.017}$	$0.965_{\pm 0.011}$	$0.920_{\pm 0.019}$	$0.739_{\pm 0.016}$	$0.869_{\pm 0.014}$
Argali Sheep	$0.868_{\pm 0.061}$	$0.987_{\pm 0.019}$	$0.934_{\pm 0.093}$	N/A	$0.917_{\pm 0.012}$
Bison	$0.713_{\pm 0.020}$	$0.945_{\pm 0.016}$	$0.783_{\pm 0.023}$	$0.501_{\pm 0.111}$	$0.769_{\pm 0.008}$
Buffalo	$0.825_{\pm 0.050}$	$0.973_{\pm 0.026}$	$0.899_{\pm 0.049}$	$0.113_{\pm 0.796}$	$0.837_{\pm 0.043}$
Cow	$0.764_{\pm 0.023}$	$0.993_{\pm 0.009}$	$0.828_{\pm 0.047}$	$0.663_{\pm 0.052}$	$0.770_{\pm 0.025}$
Sheep	$0.749_{\pm 0.014}$	$0.950_{\pm 0.022}$	$0.834_{\pm 0.025}$	$0.591_{\pm 0.214}$	$0.754_{\pm 0.015}$
Chihuahua	$0.760_{\pm 0.021}$	$0.982_{\pm 0.025}$	$0.851_{\pm 0.042}$	N/A	$0.760_{\pm 0.021}$
Collie	$0.775_{\pm 0.026}$	$0.973_{\pm 0.021}$	$0.877_{\pm 0.053}$	N/A	$0.775_{\pm 0.026}$
Dalmatian	$0.759_{\pm 0.038}$	$0.960_{\pm 0.030}$	$0.776_{\pm 0.050}$	N/A	$0.759_{\pm 0.038}$
Dog	$0.801_{\pm 0.020}$	$0.980_{\pm 0.014}$	$0.857_{\pm 0.040}$	$0.082_{\pm 0.765}$	$0.809_{\pm 0.013}$
Fox	$0.776_{\pm 0.009}$	$0.943_{\pm 0.009}$	$0.850_{\pm 0.030}$	N/A	$0.776_{\pm 0.009}$
German Shepherd	$0.777_{\pm 0.045}$	$0.958_{\pm 0.011}$	$0.864_{\pm 0.063}$	N/A	$0.777_{\pm 0.045}$
Wolf	$0.776_{\pm 0.025}$	$0.973_{\pm 0.019}$	$0.873_{\pm 0.028}$	N/A	$0.776_{\pm 0.025}$
Beaver	$0.564_{\pm 0.046}$	$0.868_{\pm 0.047}$	$0.623_{\pm 0.128}$	$0.585_{\pm 0.300}$	$0.565_{\pm 0.045}$
Alouatta	$0.764_{\pm 0.034}$	$1.000_{\pm 0.000}$	$0.922_{\pm 0.061}$	N/A	$0.764_{\pm 0.034}$
Monkey	$0.671_{\pm 0.034}$	$0.927_{\pm 0.024}$	$0.750_{\pm 0.074}$	N/A	$0.671_{\pm 0.034}$
Noisy Night Monkey	$0.755_{\pm 0.049}$	$0.973_{\pm 0.019}$	$0.847_{\pm 0.081}$	N/A	$0.759_{\pm 0.055}$
Spider Monkey	$0.573_{\pm 0.044}$	$0.871_{\pm 0.036}$	$0.620_{\pm 0.059}$	$0.267_{\pm 0.899}$	$0.571_{\pm 0.045}$
Uakari	$0.809_{\pm 0.026}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	N/A	$0.809_{\pm 0.026}$
Deer	$0.798_{\pm 0.019}$	$0.963_{\pm 0.024}$	$0.848_{\pm 0.022}$	$0.133_{\pm 0.805}$	$0.815_{\pm 0.006}$
Moose	$0.764_{\pm 0.015}$	$0.982_{\pm 0.013}$	$0.823_{\pm 0.025}$	$0.845_{\pm 0.103}$	$0.758_{\pm 0.014}$
Hamster	$0.595_{\pm 0.036}$	$0.875_{\pm 0.044}$	$0.628_{\pm 0.078}$	$0.288_{\pm 0.133}$	$0.626_{\pm 0.040}$
Elephant	$0.710_{\pm 0.030}$	$0.913_{\pm 0.048}$	$0.800_{\pm 0.049}$	N/A	$0.709_{\pm 0.031}$
Horse	$0.770_{\pm 0.043}$	$0.941_{\pm 0.031}$	$0.826_{\pm 0.069}$	$0.068_{\pm 0.758}$	$0.779_{\pm 0.038}$
Zebra	$0.748_{\pm 0.024}$	$0.910_{\pm 0.017}$	$0.760_{\pm 0.054}$	N/A	$0.761_{\pm 0.016}$
Bobcat	$0.765_{\pm 0.043}$	$0.983_{\pm 0.024}$	$0.845_{\pm 0.022}$	N/A	$0.768_{\pm 0.046}$
Cat	$0.697_{\pm 0.052}$	$0.962_{\pm 0.008}$	$0.807_{\pm 0.086}$	N/A	$0.697_{\pm 0.052}$
Cheetah	$0.756_{\pm 0.047}$	$0.942_{\pm 0.011}$	$0.821_{\pm 0.093}$	N/A	$0.785_{\pm 0.036}$
Jaguar	$0.793_{\pm 0.041}$	$0.984_{\pm 0.011}$	$0.866_{\pm 0.080}$	$0.638_{\pm 0.195}$	$0.809_{\pm 0.029}$
King Cheetah	$0.854_{\pm 0.085}$	$0.944_{\pm 0.079}$	$0.944_{\pm 0.079}$	N/A	$0.854_{\pm 0.085}$
Leopard	$0.760_{\pm 0.005}$	$0.983_{\pm 0.024}$	$0.850_{\pm 0.017}$	N/A	$0.760_{\pm 0.005}$
Lion	$0.716_{\pm 0.022}$	$0.922_{\pm 0.008}$	$0.787_{\pm 0.016}$	N/A	$0.716_{\pm 0.022}$
Panther	$0.794_{\pm 0.037}$	$0.983_{\pm 0.024}$	$0.914_{\pm 0.037}$	$0.184_{\pm 0.837}$	$0.796_{\pm 0.038}$
Persian Cat	$0.592_{\pm 0.036}$	$0.930_{\pm 0.016}$	$0.640_{\pm 0.079}$	N/A	$0.592_{\pm 0.036}$
Siamese Cat	$0.710_{\pm 0.053}$	$0.952_{\pm 0.037}$	$0.804_{\pm 0.054}$	N/A	$0.710_{\pm 0.053}$
Snow Leopard	$0.860_{\pm 0.038}$	$1.000_{\pm 0.000}$	$0.980_{\pm 0.029}$	$0.217_{\pm 0.866}$	$0.866_{\pm 0.035}$
Tiger	$0.810_{\pm 0.014}$	$0.979_{\pm 0.015}$	$0.858_{\pm 0.019}$	N/A	$0.810_{\pm 0.014}$
Giraffe	$0.799_{\pm 0.025}$	$0.949_{\pm 0.021}$	$0.840_{\pm 0.044}$	N/A	$0.805_{\pm 0.020}$
Hippo	$0.471_{\pm 0.077}$	$0.778_{\pm 0.082}$	$0.484_{\pm 0.056}$	$0.640_{\pm 0.283}$	$0.476_{\pm 0.077}$
Chimpanzee	$0.659_{\pm 0.027}$	$0.912_{\pm 0.034}$	$0.692_{\pm 0.021}$	$0.624_{\pm 0.057}$	$0.661_{\pm 0.035}$
Gorilla	$0.754_{\pm 0.014}$	$0.952_{\pm 0.037}$	$0.877_{\pm 0.013}$	N/A	$0.754_{\pm 0.014}$
Rabbit	$0.729_{\pm 0.006}$	$0.990_{\pm 0.014}$	$0.830_{\pm 0.044}$	N/A	$0.729_{\pm 0.006}$
Skunk	$0.576_{\pm 0.033}$	$0.874_{\pm 0.044}$	$0.651_{\pm 0.046}$	N/A	$0.580_{\pm 0.035}$
Mouse	$0.641_{\pm 0.017}$	$0.954_{\pm 0.033}$	$0.716_{\pm 0.039}$	$0.100_{\pm 0.778}$	$0.644_{\pm 0.018}$
Rat	$0.647_{\pm 0.040}$	$0.930_{\pm 0.044}$	$0.747_{\pm 0.040}$	$0.133_{\pm 0.818}$	$0.647_{\pm 0.044}$
Otter	$0.567_{\pm 0.012}$	$0.897_{\pm 0.020}$	$0.594_{\pm 0.017}$	N/A	$0.567_{\pm 0.012}$
Weasel	$0.694_{\pm 0.033}$	$0.935_{\pm 0.017}$	$0.781_{\pm 0.040}$	N/A	$0.694_{\pm 0.033}$
Raccoon	$0.649_{\pm 0.041}$	$0.972_{\pm 0.021}$	$0.759_{\pm 0.035}$	$0.433_{\pm 0.128}$	$0.673_{\pm 0.035}$
Rhino	$0.800_{\pm 0.011}$	$0.980_{\pm 0.000}$	$0.864_{\pm 0.031}$	N/A	$0.800_{\pm 0.011}$
Marmot	$0.792_{\pm 0.046}$	$0.967_{\pm 0.047}$	$0.917_{\pm 0.072}$	N/A	$0.792_{\pm 0.046}$
Squirrel	$0.863_{\pm 0.008}$	$1.000_{\pm 0.000}$	$0.953_{\pm 0.013}$	N/A	$0.863_{\pm 0.008}$
Pig	$0.645_{\pm 0.040}$	$0.928_{\pm 0.050}$	$0.706_{\pm 0.039}$	N/A	$0.644_{\pm 0.040}$
Black Bear	$0.685_{\pm 0.035}$	$1.000_{\pm 0.000}$	$0.853_{\pm 0.144}$	N/A	$0.685_{\pm 0.035}$
Brown Bear	$0.612_{\pm 0.010}$	$0.960_{\pm 0.014}$	$0.680_{\pm 0.042}$	$0.068_{\pm 0.755}$	$0.615_{\pm 0.008}$
Panda	$0.572_{\pm 0.028}$	$0.935_{\pm 0.006}$	$0.604_{\pm 0.091}$	N/A	$0.571_{\pm 0.029}$
Polar Bear	$0.623_{\pm 0.001}$	$0.931_{\pm 0.053}$	$0.654_{\pm 0.015}$	N/A	$0.623_{\pm 0.001}$

Table S5: Per-species results of HRNet-w32 [13] on the **56 animals** test set the SL Track at the setting of training from scratch.

species	AP	$AP_{.5}$	$AP_{.75}$	AP_M	AP_L
Antelope	$0.834_{\pm 0.019}$	$0.964_{\pm 0.011}$	$0.910_{\pm 0.013}$	$0.694_{\pm 0.026}$	$0.854_{\pm 0.017}$
Argali Sheep	$0.856_{\pm 0.087}$	$0.957_{\pm 0.061}$	$0.932_{\pm 0.096}$	N/A	$0.909_{\pm 0.014}$
Bison	$0.700_{\pm 0.041}$	$0.925_{\pm 0.033}$	$0.778_{\pm 0.034}$	$0.524_{\pm 0.138}$	$0.754_{\pm 0.025}$
Buffalo	$0.807_{\pm 0.038}$	$0.976_{\pm 0.033}$	$0.876_{\pm 0.051}$	$0.091_{\pm 0.779}$	$0.820_{\pm 0.033}$
Cow	$0.756_{\pm 0.013}$	$0.980_{\pm 0.014}$	$0.838_{\pm 0.018}$	$0.644_{\pm 0.080}$	$0.762_{\pm 0.013}$
Sheep	$0.749_{\pm 0.013}$	$0.958_{\pm 0.024}$	$0.831_{\pm 0.019}$	$0.616_{\pm 0.160}$	$0.755_{\pm 0.013}$
Chihuahua	$0.716_{\pm 0.029}$	$0.949_{\pm 0.002}$	$0.776_{\pm 0.071}$	N/A	$0.718_{\pm 0.026}$
Collie	$0.763_{\pm 0.010}$	$0.983_{\pm 0.012}$	$0.840_{\pm 0.035}$	N/A	$0.763_{\pm 0.010}$
Dalmatian	$0.723_{\pm 0.041}$	$0.916_{\pm 0.025}$	$0.731_{\pm 0.055}$	N/A	$0.723_{\pm 0.041}$
Dog	$0.782_{\pm 0.016}$	$0.964_{\pm 0.014}$	$0.850_{\pm 0.013}$	$0.149_{\pm 0.825}$	$0.791_{\pm 0.019}$
Fox	$0.738_{\pm 0.017}$	$0.939_{\pm 0.008}$	$0.803_{\pm 0.049}$	N/A	$0.738_{\pm 0.017}$
German Shepherd	$0.733_{\pm 0.041}$	$0.945_{\pm 0.005}$	$0.840_{\pm 0.061}$	N/A	$0.734_{\pm 0.040}$
Wolf	$0.751_{\pm 0.040}$	$0.971_{\pm 0.020}$	$0.794_{\pm 0.070}$	N/A	$0.751_{\pm 0.040}$
Beaver	$0.529_{\pm 0.034}$	$0.879_{\pm 0.023}$	$0.557_{\pm 0.087}$	$0.584_{\pm 0.245}$	$0.528_{\pm 0.033}$
Alouatta	$0.725_{\pm 0.045}$	$1.000_{\pm 0.000}$	$0.876_{\pm 0.120}$	N/A	$0.725_{\pm 0.045}$
Monkey	$0.629_{\pm 0.052}$	$0.930_{\pm 0.023}$	$0.686_{\pm 0.082}$	N/A	$0.629_{\pm 0.052}$
Noisy Night Monkey	$0.721_{\pm 0.004}$	$0.958_{\pm 0.001}$	$0.758_{\pm 0.040}$	N/A	$0.725_{\pm 0.010}$
Spider Monkey	$0.535_{\pm 0.029}$	$0.830_{\pm 0.048}$	$0.588_{\pm 0.033}$	$0.233_{\pm 0.873}$	$0.533_{\pm 0.031}$
Uakari	$0.763_{\pm 0.028}$	$1.000_{\pm 0.000}$	$0.921_{\pm 0.017}$	N/A	$0.763_{\pm 0.028}$
Deer	$0.792_{\pm 0.031}$	$0.960_{\pm 0.029}$	$0.836_{\pm 0.037}$	$0.130_{\pm 0.804}$	$0.808_{\pm 0.011}$
Moose	$0.759_{\pm 0.017}$	$0.959_{\pm 0.008}$	$0.844_{\pm 0.019}$	$0.691_{\pm 0.136}$	$0.761_{\pm 0.019}$
Hamster	$0.520_{\pm 0.043}$	$0.829_{\pm 0.079}$	$0.535_{\pm 0.035}$	$0.276_{\pm 0.138}$	$0.553_{\pm 0.024}$
Elephant	$0.672_{\pm 0.038}$	$0.883_{\pm 0.031}$	$0.772_{\pm 0.051}$	N/A	$0.671_{\pm 0.040}$
Horse	$0.757_{\pm 0.015}$	$0.935_{\pm 0.022}$	$0.785_{\pm 0.008}$	N/A	$0.768_{\pm 0.012}$
Zebra	$0.725_{\pm 0.024}$	$0.872_{\pm 0.024}$	$0.752_{\pm 0.038}$	N/A	$0.733_{\pm 0.020}$
Bobcat	$0.727_{\pm 0.054}$	$0.973_{\pm 0.021}$	$0.793_{\pm 0.088}$	N/A	$0.727_{\pm 0.054}$
Cat	$0.667_{\pm 0.040}$	$0.943_{\pm 0.006}$	$0.755_{\pm 0.040}$	N/A	$0.672_{\pm 0.041}$
Cheetah	$0.718_{\pm 0.020}$	$0.956_{\pm 0.010}$	$0.751_{\pm 0.083}$	N/A	$0.742_{\pm 0.031}$
Jaguar	$0.748_{\pm 0.016}$	$0.971_{\pm 0.021}$	$0.843_{\pm 0.021}$	$0.649_{\pm 0.230}$	$0.764_{\pm 0.013}$
King Cheetah	$0.798_{\pm 0.079}$	$0.944_{\pm 0.079}$	$0.832_{\pm 0.138}$	N/A	$0.798_{\pm 0.079}$
Leopard	$0.705_{\pm 0.004}$	$0.938_{\pm 0.030}$	$0.782_{\pm 0.062}$	N/A	$0.705_{\pm 0.004}$
Lion	$0.685_{\pm 0.026}$	$0.915_{\pm 0.019}$	$0.714_{\pm 0.046}$	N/A	$0.685_{\pm 0.026}$
Panther	$0.748_{\pm 0.041}$	$0.983_{\pm 0.024}$	$0.798_{\pm 0.044}$	$0.217_{\pm 0.863}$	$0.748_{\pm 0.040}$
Persian Cat	$0.559_{\pm 0.046}$	$0.858_{\pm 0.015}$	$0.535_{\pm 0.070}$	N/A	$0.559_{\pm 0.046}$
Siamese Cat	$0.677_{\pm 0.038}$	$0.965_{\pm 0.025}$	$0.713_{\pm 0.063}$	N/A	$0.677_{\pm 0.038}$
Snow Leopard	$0.826_{\pm 0.058}$	$1.000_{\pm 0.000}$	$0.866_{\pm 0.077}$	$0.167_{\pm 0.850}$	$0.831_{\pm 0.048}$
Tiger	$0.764_{\pm 0.021}$	$0.962_{\pm 0.034}$	$0.820_{\pm 0.008}$	N/A	$0.764_{\pm 0.021}$
Giraffe	$0.783_{\pm 0.015}$	$0.949_{\pm 0.015}$	$0.815_{\pm 0.027}$	N/A	$0.790_{\pm 0.012}$
Hippo	$0.383_{\pm 0.062}$	$0.659_{\pm 0.070}$	$0.383_{\pm 0.058}$	$0.615_{\pm 0.245}$	$0.384_{\pm 0.057}$
Chimpanzee	$0.625_{\pm 0.031}$	$0.944_{\pm 0.026}$	$0.709_{\pm 0.051}$	$0.547_{\pm 0.114}$	$0.629_{\pm 0.038}$
Gorilla	$0.736_{\pm 0.029}$	$0.942_{\pm 0.036}$	$0.836_{\pm 0.058}$	N/A	$0.736_{\pm 0.029}$
Rabbit	$0.712_{\pm 0.032}$	$0.979_{\pm 0.015}$	$0.791_{\pm 0.031}$	N/A	$0.712_{\pm 0.032}$
Skunk	$0.514_{\pm 0.032}$	$0.832_{\pm 0.013}$	$0.566_{\pm 0.089}$	N/A	$0.518_{\pm 0.032}$
Mouse	$0.608_{\pm 0.032}$	$0.928_{\pm 0.030}$	$0.657_{\pm 0.090}$	$0.124_{\pm 0.795}$	$0.607_{\pm 0.032}$
Rat	$0.588_{\pm 0.079}$	$0.935_{\pm 0.058}$	$0.623_{\pm 0.128}$	$0.133_{\pm 0.806}$	$0.586_{\pm 0.083}$
Otter	$0.537_{\pm 0.021}$	$0.889_{\pm 0.037}$	$0.516_{\pm 0.048}$	N/A	$0.537_{\pm 0.021}$
Weasel	$0.654_{\pm 0.027}$	$0.917_{\pm 0.012}$	$0.715_{\pm 0.019}$	N/A	$0.656_{\pm 0.028}$
Raccoon	$0.596_{\pm 0.014}$	$0.935_{\pm 0.017}$	$0.588_{\pm 0.063}$	$0.452_{\pm 0.178}$	$0.612_{\pm 0.010}$
Rhino	$0.773_{\pm 0.043}$	$0.945_{\pm 0.033}$	$0.854_{\pm 0.076}$	N/A	$0.773_{\pm 0.043}$
Marmot	$0.770_{\pm 0.037}$	$1.000_{\pm 0.000}$	$0.836_{\pm 0.064}$	N/A	$0.770_{\pm 0.037}$
Squirrel	$0.840_{\pm 0.019}$	$1.000_{\pm 0.000}$	$0.951_{\pm 0.037}$	N/A	$0.840_{\pm 0.019}$
Pig	$0.616_{\pm 0.060}$	$0.902_{\pm 0.062}$	$0.712_{\pm 0.095}$	N/A	$0.615_{\pm 0.059}$
Black Bear	$0.682_{\pm 0.066}$	$0.967_{\pm 0.047}$	$0.807_{\pm 0.099}$	N/A	$0.686_{\pm 0.061}$
Brown Bear	$0.589_{\pm 0.022}$	$0.973_{\pm 0.005}$	$0.621_{\pm 0.050}$	$0.101_{\pm 0.779}$	$0.589_{\pm 0.022}$
Panda	$0.525_{\pm 0.033}$	$0.921_{\pm 0.024}$	$0.498_{\pm 0.081}$	N/A	$0.524_{\pm 0.033}$
Polar Bear	$0.599_{\pm 0.022}$	$0.912_{\pm 0.037}$	$0.639_{\pm 0.015}$	N/A	$0.599_{\pm 0.022}$

A.2 Motivation

1. For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

A1: AP-10K is created to facilitate research in the area of animal pose estimation. It is important to study several challenging questions in the context of more training data from diverse species are available, such as 1) how about the performance of different representative human pose models on the animal pose estimation task? 2) will the representation ability of a deep model benefit from training on a large-scale dataset with diverse species? 3) how about the impact of pretraining, *e.g.*, on the ImageNet dataset [5] or human pose estimation dataset [9], in the context of the large-scale of dataset with diverse species? and 4) how about the intra- and inter-family generalization ability of a model trained using data from specific species or family? However, previous datasets for animal pose estimation contain limited number of animal species. Therefore, it is impossible to study these questions using existing datasets as they contains at most 5 species, which is far from enough to get sound conclusion. By contrast, AP-10K has 23 family and 54 species and thus can help researchers to study these questions.

2. Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

A2: AP-10K is created by the authors as well as some volunteer graduate students from Xidian University, including Jiarui Fan, Ziming Bai, Jinglong Zhao, Yan Liu, Yi Liu, Jiaming Li, Hanbo Sun, Chong Guo, Junwei Duan, and Xinyu Wang.

3. Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

A3: The creation of the dataset is founded by the Innovation Capability Support Program of Shaanxi under the grant of Program No.2021TD-05 and the National Natural Science Foundation of China under the grant of No.62133012, No.61936006.

A.3 Composition

1. What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances(e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

A1: AP-10K is comprised of images covering 54 animal species, which are categorized following the taxonomic rank, *i.e.*, family and species. For each animal instance, its 17 keypoint annotations are provided, including Left Eye, Right Eye, Nose, Neck, Root of Tail, Left Shoulder, Left Elbow, Left Front Paw, Right Shoulder, Right Elbow, Right Front Paw, Left Hip, Left Knee, Left Back Paw, Right Hip, Right Knee, Right Back Paw.

2. How many instances are there in total (of each type, if appropriate)?

A2: The AP-10K dataset contains 10,015 images and 13,028 instances with keypoint annotations. Besides, AP-10K contains extra 49,643 images, where each image is only annotated with family and species labels, without keypoint annotations.

3. Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

A3: AP-10K is a real-world sample of animals, including information about their poses. It is the largest dataset in the area of animal pose estimation, *e.g.*, 10× more species compared with other previous datasets. Due to the diversity of real-world animal species, it is impossible to cover all instances of wild animals on a single dataset. The AP-10K dataset contains 23 typical families and 54 species, which follows the taxonomic rank, providing a more diverse set of instances than ever before and will facilitate further studies of animal pose estimation. Biological proximity exploited in AP-10K may be used to help the model extend to more species in the wild.

4. What data does each instance consist of? Raw data (e.g., unprocessed text or images)or features? In either case, please provide a description.

A4: Each instance consists of one animal with its location (bounding box), family and species labels, keypoint annotations, and the unprocessed image data.

5. Is there a label or target associated with each instance? If so, please provide a description.

A5: Yes. Each target is associated with labels following the COCO-style, which contain the instance id, image id, category information (family and species), the box area, whether is crowed or not, the number of keyponts, detailed keypoints information (location and category), as well as the background category.

6. Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

A6: Yes. Some instances may not have the complete keypoint annotation due to occlusion, blur, or small scale, similar to those instances in the COCO human pose dataset. Therefore, the location and visibility of these keypoints are marked zero following the COCO-style.

7. Are relationships between individual instances made explicit (e.g., users movie ratings, social network links)? If so, please describe how these relationships are made explicit.

A7: Yes. The instances' information are stored in the COCO-style, where the relationship between instances, *e.g.*, whether two instances belong to the same category or are on the same image, can be queried using the COCO APIs.

8. Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

A8: Yes. We randomly split the dataset into disjoint train, validation, and test sets following the ratio of 7:1:2. The splits are done within each species to keep the origin distribution.

9. Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

A9: Although we have carefully double-check the annotations, there may be some inaccurate keypoint annotations, *e.g.*, small drifts in the annotation of keypoint locations. To figure out the error rate of AP-10K annotations, we re-examine the annotations, and the results are available in Table S6.

Table S6: Error rate in AP-10K.

Number of Images	Total keypoints	Mislabeled keypoints	Error rate
10015	130142	15	0.012%

10. Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

A10: The AP-10K is comprised of the publicly available datasets, including African Wildlife [6], Animal-Pose Dataset [3], Animal Image Dataset(DOG, CAT and PANDA) [11], Endangered Animals [8], IUCN Animals Dataset [2], Animals with Attributes 2 [14], Animals-5 [12], Animals-10 [1], and Wild Cats [16]. These datasets are publicly avaiable and can be downloaded from their websites, *e.g.*, the Animals-10 dataset follows the GPL v2 license while Endangered Animals follows the CC-BY-4.0 license. We appreciate the significant contribution of the authors to the research community. Table S7 shows the number of images contained in the each dataset that contribute to the AP-10K dataset.

Table S7: The source dataset adopted in AP-10K.

Source Dataset	Number of Images
African Wildlife [6]	1454
Animal Image Dataset(DOG, CAT and PANDA) [11]	2753
Animal-Pose [3]	970
Animals with Attributes 2 [14]	31701
Animals-5 [12]	5110
Animals-10 [1]	14029
Endangered Animals [8]	284
IUCN Animals Dataset [2]	1689
Wild Cats [16]	1558
Internet	110

11. Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctorpatient confidentiality, data that includes the content of individuals non-public communications)? If so, please provide a description.

A11: No.

12. Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

A12: No.

A.4 Collection Process

1. How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

A1: The data associated with each instance are directly observable, as they are stored as the common COCO format and can be easily viewed via the COCO APIs.

2. What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated?

A2: The images in AP-10K come from dataset publicly available datasets described above, which can be directly downloaded from their websites.

3. If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

A3: No.

4. Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

A4: The first author of this paper.

5. Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

A5: It took about 1 day to collect the data and about 3 months to complete organization and annotation, as each participant labelled the bonding boxes and keypoints about one hour per workday.

A.5 Preprocessing/cleaning/labeling

1. Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing

of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section.

A1: We remove replicated images by using aHash [4] to detect similar images and manually checking. Then, images with heavy occlusion and logos are removed manually. These images are categorized into family and species, which are double checked by the annotators to ensure the image quality of AP-10K. Next, the annotators after training start to annotate the images.

During training, annotators first learned about the physiognomy, body structure and distribution of keypoints of the animals. Then, five images of each species were presented to annotators to annotate keypoints, which were used to assess their annotation quality. Annotators with good annotation quality were further trained on how to deal with the partial absence of the body due to occlusion and were involved in the subsequent annotation process. Annotators were asked to annotate all visible keypoints. For the occluded keypoints, they were asked to annotate keypoints whose location they could estimate based on body plan, pose, and the symmetry property of the body, where the length of occluded limbs or the location of occluded keypoints could be inferred from the visible limbs or keypoints. Other keypoints were left unlabeled. As shown in Figure 1, the right limb of the polar bear on the right could not be estimated accurately, so the annotator did not annotate it, while the right limb of the panda was annotated because it could be easily estimated from their pose and the symmetry property of the body.

To guarantee the annotation quality, we have adopted a sequential labeling strategy. Three rounds of cross-check and correction are conducted with both manual check and automatic check (according to specific rules, *e.g.*, keypoints belonging to an instance are in the same bounding box) to reduce the possibility of mislabeling.

2. Was the raw data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the raw data.

A2: No.

3. Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.

A3: We use the open source labelling tool.

A.6 Uses

1. Has the dataset been used for any tasks already? If so, please provide a description.

A1: No.

2. Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

A2: N/A.

3. What (other) tasks could the dataset be used for?

A3: AP-10K can be used for the research of animal pose estimation. Besides, it can also be used for specific machine learning topics such as few-shot learning, domain generalization, self-supervised learning. Please see the Discussion part in the paper.

4. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?

A4: No.

5. Are there tasks for which the dataset should not be used? If so, please provide a description.

A5: No.

A.7 Distribution

1. Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

A1: Yes. The dataset will be made publicly available to the research community.

2. How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

A2: It will be publicly available on the project website at GitHub.

3. When will the dataset be distributed?

A3: The dataset will be distributed once the paper is accepted after peer-review.

4. Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

A4: It will be distributed under the MIT licence.

5. Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

A5: No.

6. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

A6: No.

A.8 Maintenance

1. Who will be supporting/hosting/maintaining the dataset?

A1: The authors.

2. How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

A2: They can be contacted via email available on the project website.

3. Is there an erratum? If so, please provide a link or other access point.

A3: No.

4. Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)?

A4: No. We have carefully double checked the annotations to reduce the labeling errors. There may be a very few of labeling errors, which can be treated as noise.

5. Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users.

A5: N/A.

6. If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description.

A6: N/A.

Figure S1: representative examples of failures of the pose estimation in our AP-10K.

References

[1] C. Alessio. Animals 10. https://www.kaggle.com/alessiocorrado99/animals10, 2019.

[2] antoreepjana. Iucn animals dataset. https://www.kaggle.com/antoreepjana/iucn-animals-dataset, 2021.

[3] J. Cao, H. Tang, H.-S. Fang, X. Shen, C. Lu, and Y.-W. Tai. Cross-domain adaptation for animal pose estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[4] B. P. CLi Weng. A secure perceptual hash algorithm for image content authentication. 2011.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] B. Ferreira. African wildlife. https://www.kaggle.com/biancaferreira/african-wildlife, 2021.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] S. Jamil. Endangered animals. https://www.kaggle.com/sonain/endangered-animals, 2020.

[9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[10] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016.

[11] A. Saxena. Animal image dataset(dog, cat and panda). https://www.kaggle.com/ashishsaxena2209/animal-image-datasetdog-cat-and-panda, 2019.

[12] Y. V. Trivedi. Animals 5. https://www.kaggle.com/ytrivedi1/animals-5, 2020.

[13] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[14] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata. Zero-shot learninga comprehensive evaluation of the good, the bad and the ugly. *IEEE transactions on pattern analysis and machine intelligence*, 41(9):2251–2265, 2018.

[15] B. Xiao, H. Wu, and Y. Wei. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 466–481, 2018.

[16] E. ahovi. wild cats. https://www.kaggle.com/enisahovi/cats-projekat-4, 2020.

Figure S2: Some visual examples of animal species and annotations in our AP-10K.

Figure S3: Some visual examples of animal species and annotations in our AP-10K.

Figure S4: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Antelope, Argali Sheep, Deer, Moose, and Giraffe, respectively.

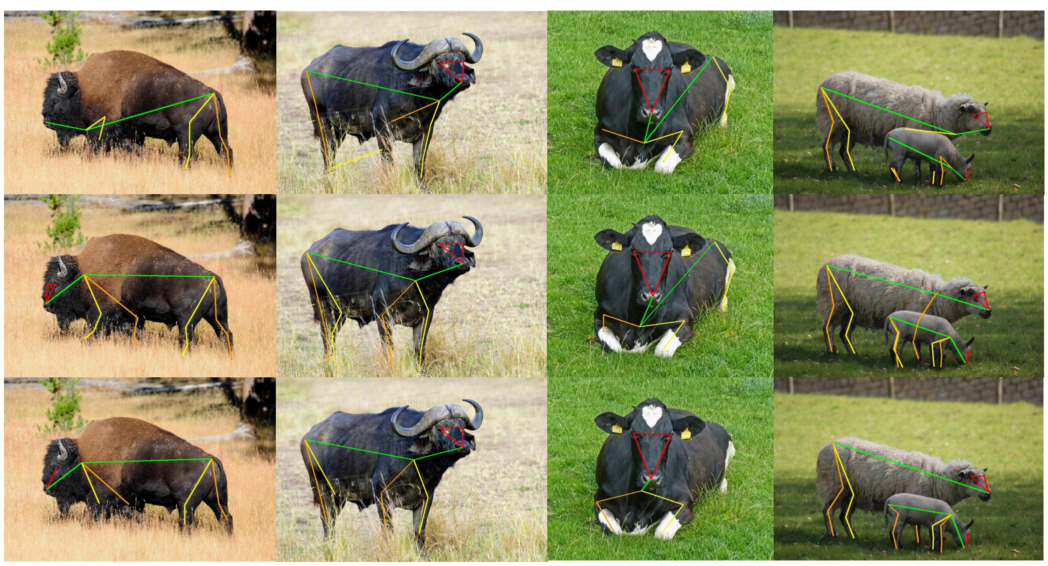

Figure S5: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Bison, Buffalo, Cow, and Sheep, respectively.

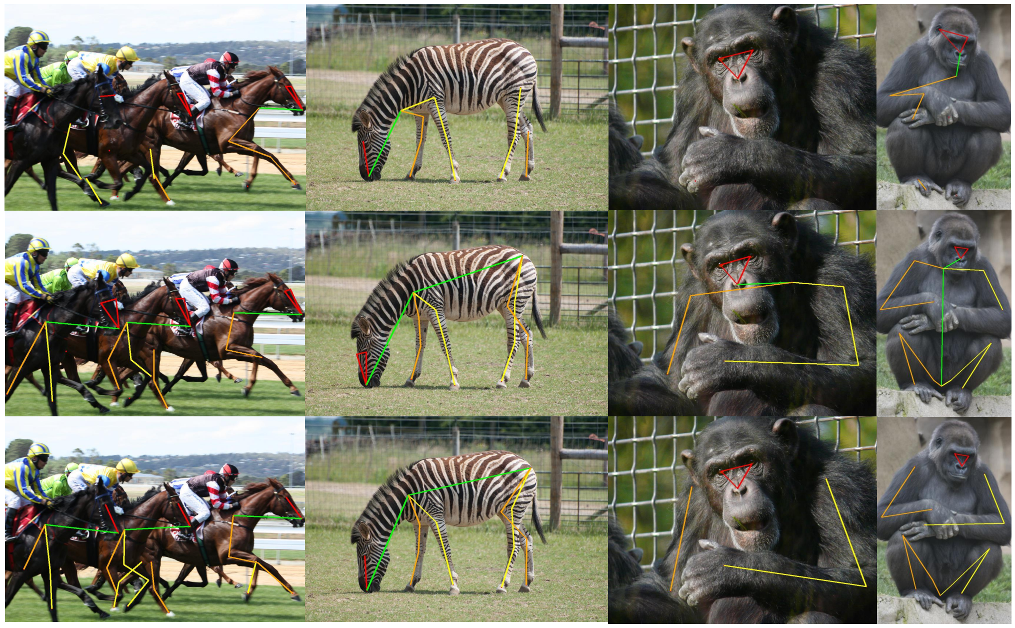

Figure S6: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Horse, Zebra, Chimpanzee, and Gorilla, respectively.

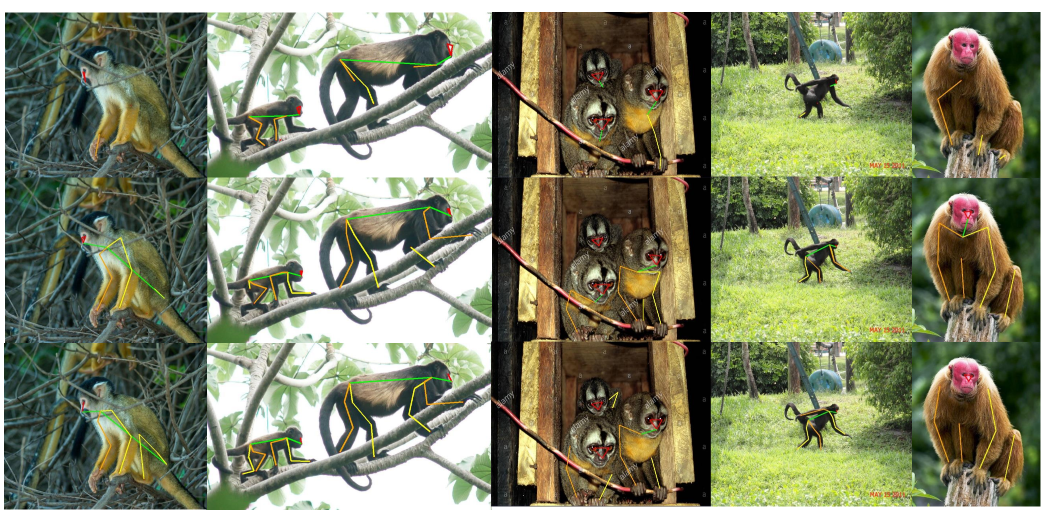

Figure S7: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Monkey, Alouatta, Noisy Night Monkey, Spider Monkey, and Uakari, respectively.

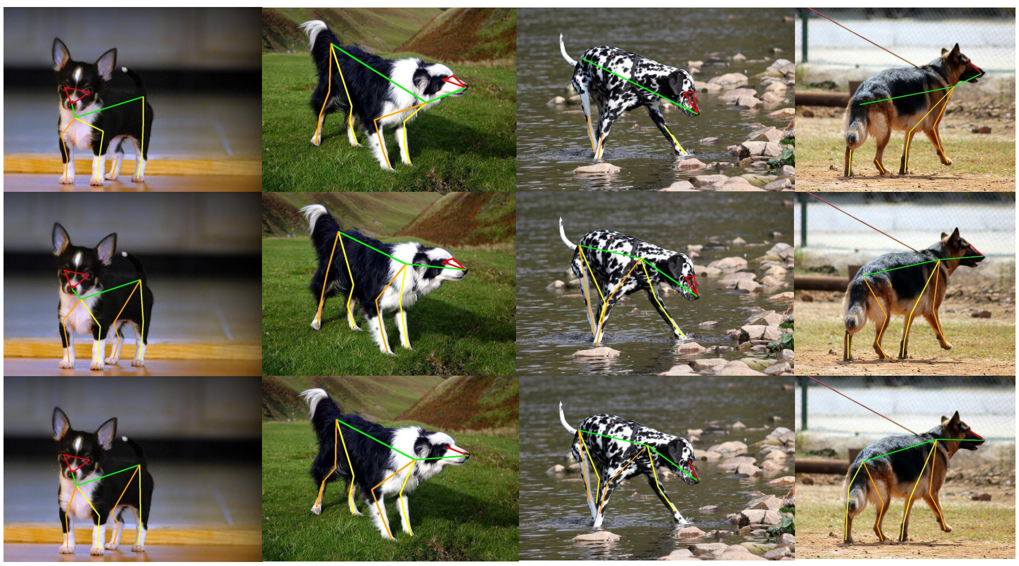

Figure S8: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Chihuauha, Collie, Dalmatian, and German Shepherd, respectively.

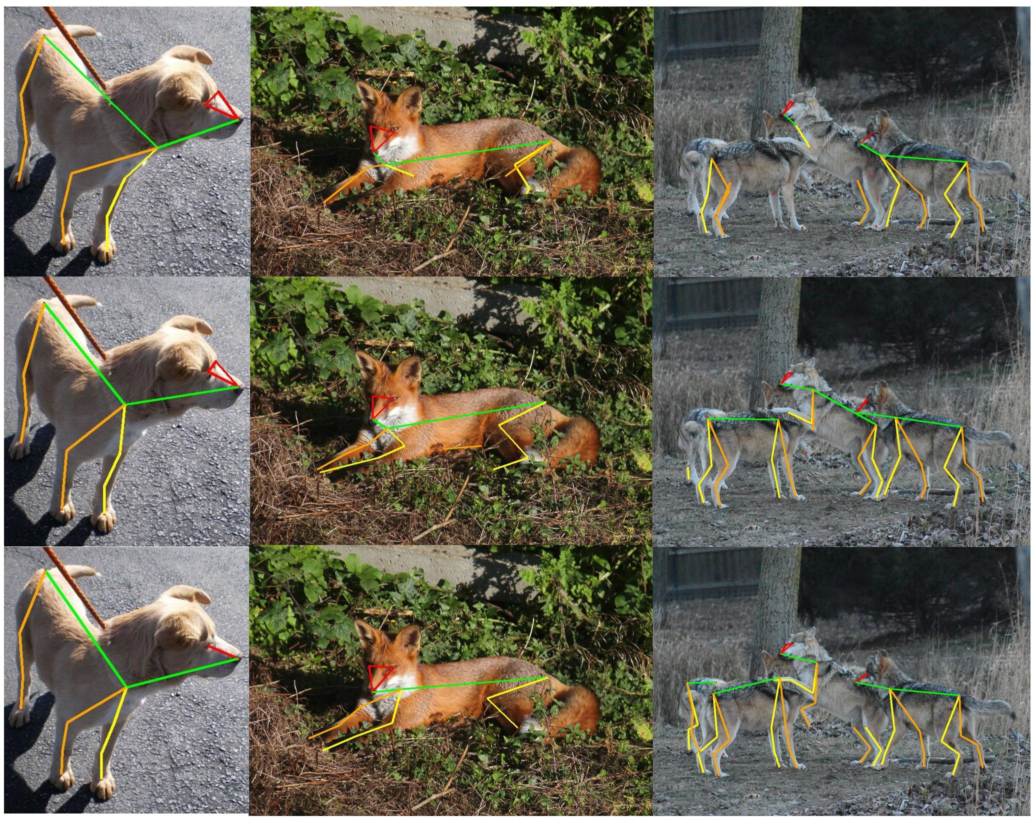

Figure S9: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Dog, Fox, and Wolf, respectively.

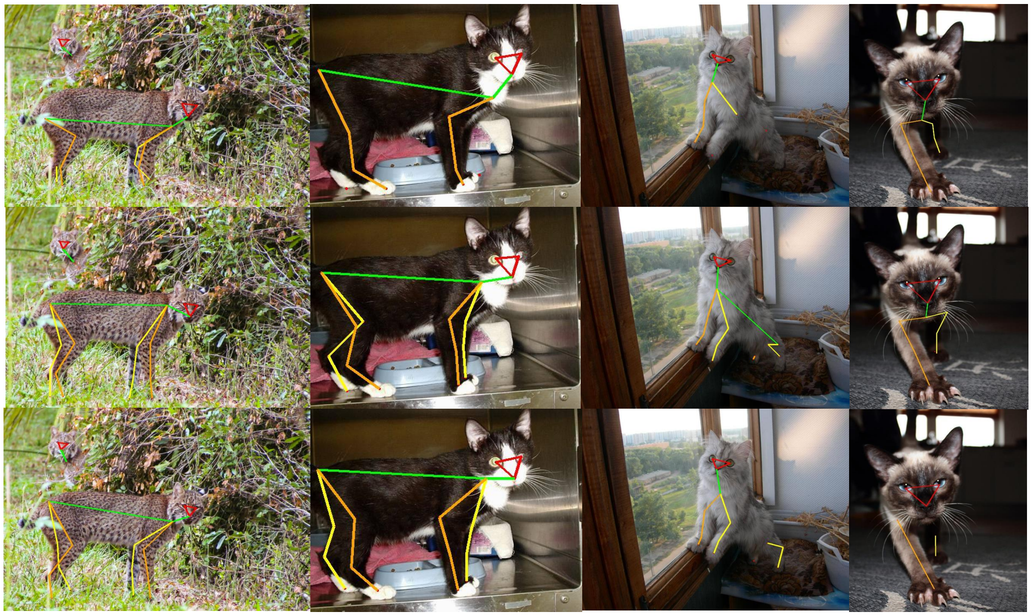

Figure S10: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Bobcat, Cat, Persian Cat, and Siamese Cat, respectively.

Figure S11: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Cheetah, King Cheetah, Leopard, and Panther, respectively.

Figure S12: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Snow Leopard, Lion, Jaguar, and Tiger, respectively.

Figure S13: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Black Bear, Brown Bear, Panda, and Polar Bear, respectively.

Figure S14: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. Elephant, Hippo, Rhino, and Pig, respectively.

Figure S15: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Skunk, Raccoon, Squirrel, and Marmot, respectively.

Figure S16: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row. These animals are Beaver, Otter, and Weasel, respectively.

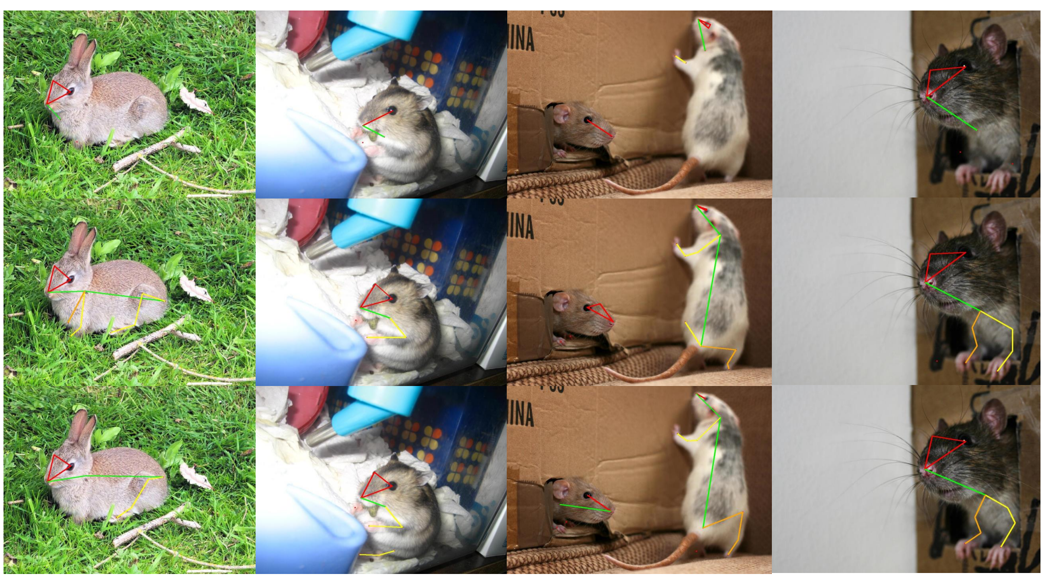

Figure S17: More qualitative results of HRNet-w32 trained on the Animal Pose dataset [3] (the first row) and our AP-10K dataet (the second row). The ground truth poses are shown in the last row.These animals are Rabbit, Hamster, Mouse, and Rat, respectively.

Figure S18: The results on video for animals in the wild.

Figure S19: The results on video for animals in the wild.

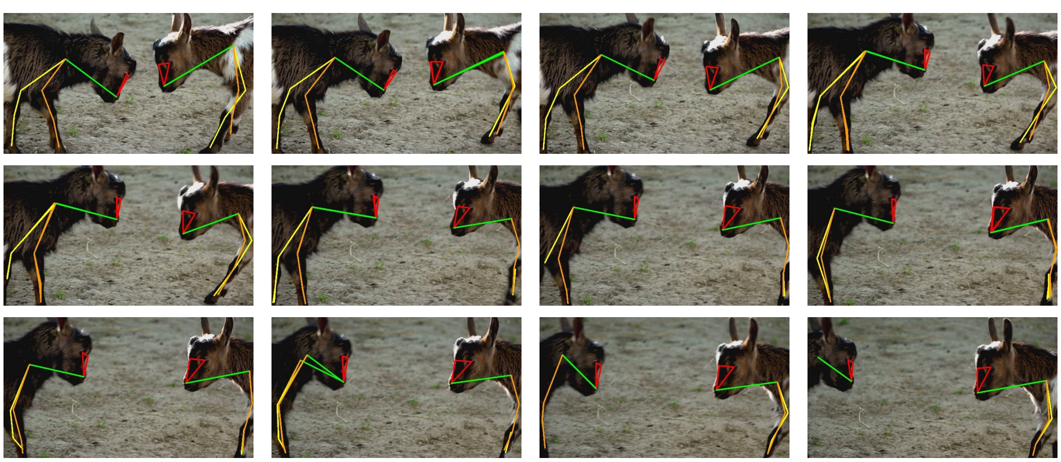

Figure S20: The results on video for animals in the wild.

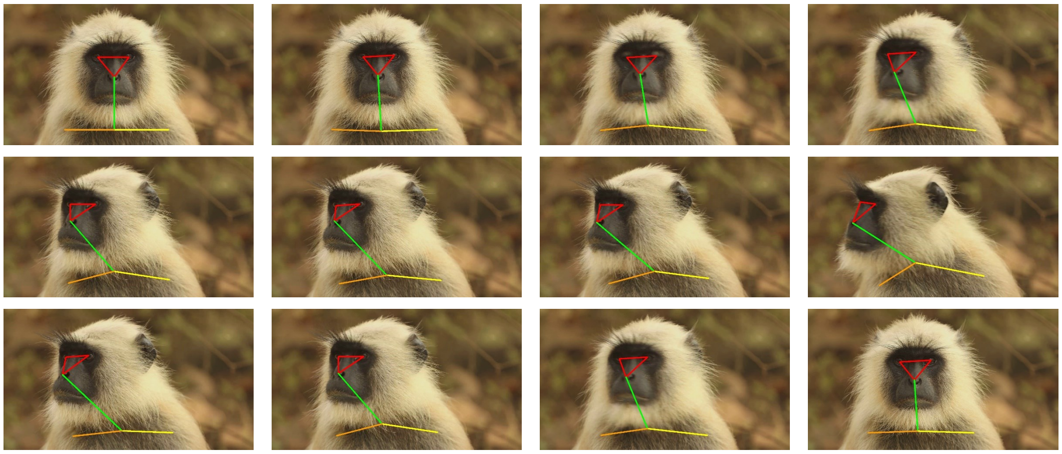

Figure S21: The results on video for animals in the wild.

Figure S22: The results on video for animals in the wild (different scenes).