LoTE-Animal: A Long Time-span Dataset for Endangered Animal Behavior Understanding
(Supplementary material 2: More Information and Results)

Dan Liu[1]*, Jin Hou[2,3]*, Shaoli Huang[4]*, Jing Liu[1], Yuxin He[5], Bochuan Zheng[2]†, Jifeng Ning[1]†, Jindong Zhang[2]†
[1]Northwest A&F University, [2]China West Normal University, [3]Beijing Normal University,
[4]Tencent AI-Lab, [5]North Sichuan Medical College
{dliu, iliujing}@nwafu.edu.cn, {houj815, heyuxin_nsmc, zhangjd224}@163.com,
shaolihuang@tencent.com, zhengbc@vip.163.com, njf@nwsuaf.edu.cn

1. More information about the dataset

In this section, we provide more statistical information about the dataset, such as the year, season, and day/night distribution.

1.1. Year span

The data were collected from 2009 to 2021, covering a time span of approximately 12 years. The total number of camera working days was in the thousands. The image resolutions varied from (4000,3000), (1920,1080), to (640,480). Infrared trap cameras are particularly suitable for protecting endangered species with narrow distribution ranges, low density, and small numbers, as shown in Fig. 1. The cameras were mainly set up in areas where animals nest, traverse, drink water, on ridges, and where food resources are abundant. During the installation of infrared cameras, we mainly selected small habitats where animals frequently gather based on their activity traces such as feces, footprints, and food traces. We used infrared trap cameras to record the activities of large and medium-sized or nocturnal rare wildlife. These images truly record the lives of wild animals that are unknown to ordinary people, many of which are first disclosed in terms of behavior and appearance, with extremely valuable scientific research value.

1.2. Season distribution

The Wolong National Nature Reserve is located in the Qinghai-Tibet Plateau climate region, mainly divided into three ecological seasons: November of the current year to March of the next year is the winter season of Wolong, with snow accumulation and high humidity; April to June is the spring season of Wolong, with temperature rising and rainfall increasing; July to October is the summer and autumn season of Wolong, with significant increases in rainfall, solar radiation intensity, and temperature.

LoTE-animal dataset collected images from different seasons, including spring, summer, autumn, and winter, with different vegetation cover and solar radiation intensity in each season. Each image in the dataset is labeled with the season, providing information about the seasonal distribution, activities, and population numbers of certain species. We calculated the number of photos collected in each season and presented them in Fig. 2.

1.3. Day-Night distribution

Furthermore, the wild subset of the LoTE-animal dataset is labeled with day and night, which provides information on the diurnal and nocturnal activity frequencies of certain species, even analyzing the activity changes from diurnal to nocturnal or from diurnal and nocturnal hunting to nocturnal hunting. We conducted statistics on the overall distribution of species' diurnal and nocturnal activities based on the camera's captured images, as shown in Fig. 3.

2. More confusion matrix results

In this section, we visualize the confusion matrix of TOOD of object detection on wild subset, web subset and cross subset, which are shown in Fig. 4, Fig. 5 and Fig. 6 respectively.

We also visualized the confusion matrix of instance segmentation Queryinst on wild subset, web subset and cross subset, which are shown in Fig. 7, Fig. 8 and Fig. 9 respectively.

3. More qualitative visualization results

In this section, we provide more qualitative visualization results on the performance of visual tasks in wild scenes.

*These authors contribute equally to this work.
†Corresponding authors: Bochuan Zheng, Jifeng Ning, Jindong Zhang.

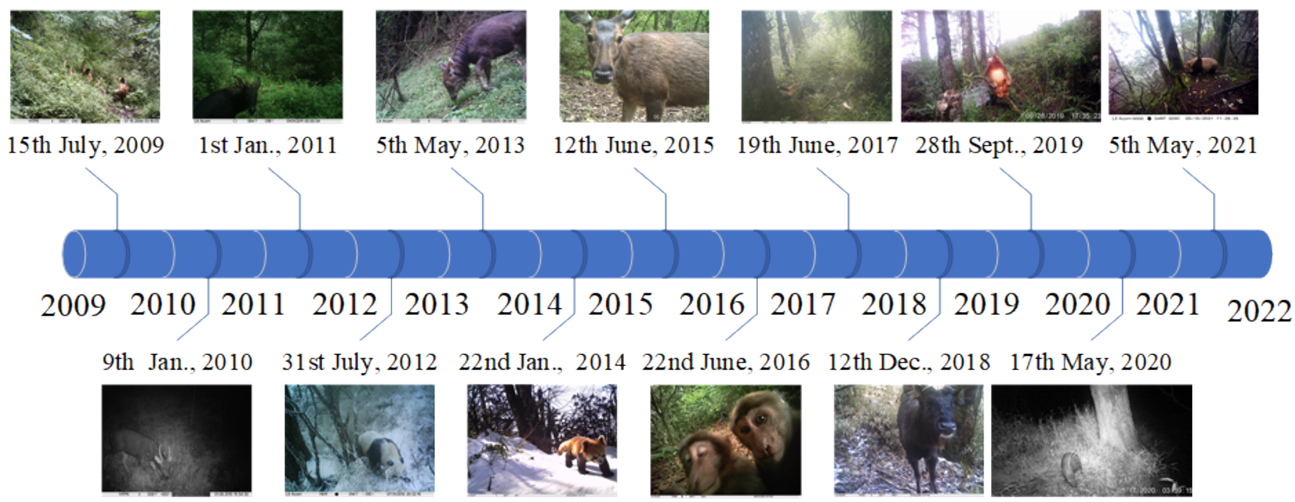

Figure 1: Year span.

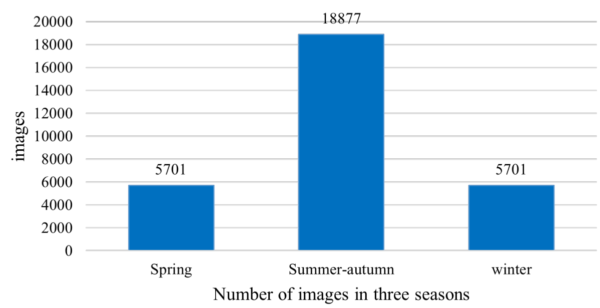

Figure 2: Season distribution.

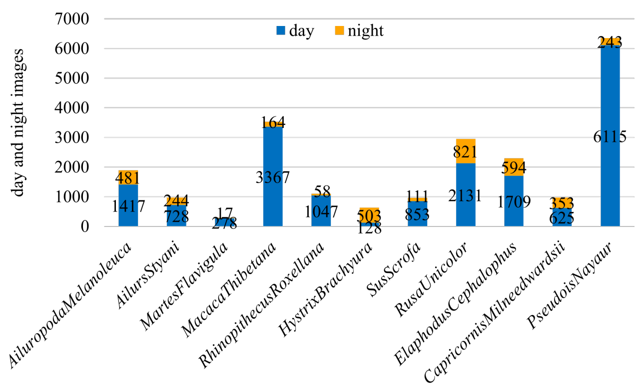

Figure 3: Day-Night distribution.

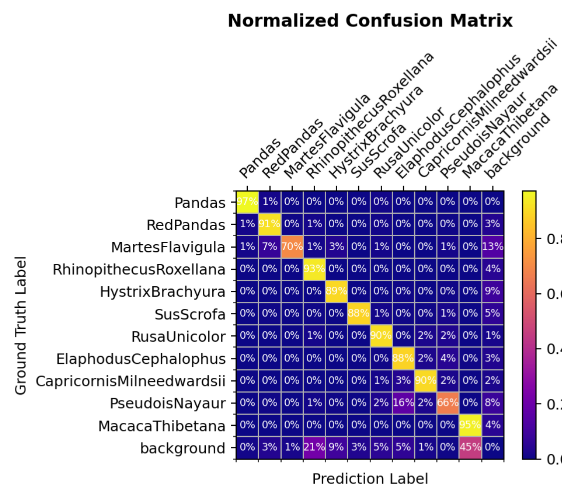

Figure 4: Tood on wild subset.

3.1. Object detection

We analyze the visual results of wild animal object detection and explore the reasons for poor prediction performance. We found that problems such as distant and small targets, occlusion, blurring, and extreme similarity with the background all negatively affect the performance of object detection. Additionally, we also found other issues, such as animals occupying too much space, ground truth being large but predicted box being small, one animal being detected as two, and multiple animals being predicted as one detection box, as well as false detections.

The visualization of these questions are as follows. Note: (Green boxes: GT (Ground Truth), red boxes: DT (Detect Truth)).

1. Remote, small animals

Remote, small animals are easy to miss detection, and most of them are monkeys, followed by distant blue sheep(*Pseudois Nayaur*) and young wild boar (*Sus Scrofa*), as shown in Fig. 10.

2. Occlusion problem

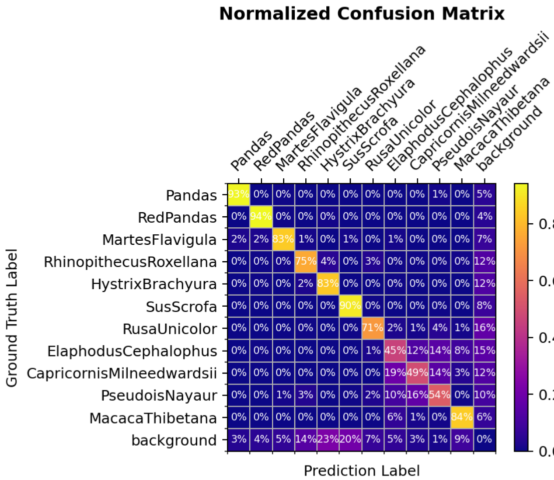

Figure 5: Tood on web subset.

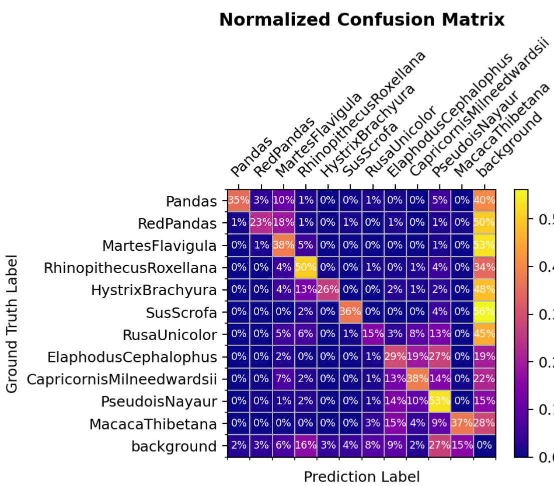

Figure 6: Tood on cross subset.

(1) The first type is animal occlusion, such as mother holding baby, mounting, animals too close to each other, as shown in Fig. 11.

(2) The second type is environmental occlusion, such as the occlusion of trees and grass, as shown in Fig. 12.

3. Blurry image problem

(1) The first type is the illumination, which includes dark shadow at night, overexposure during the day and silhouette on cloudy day, as shown in Fig. 13.

(2) The second type is foggy weather, resulting in blurred images, as shown in Fig. 14.

(3) The third type is that the image itself has a low pixel, as shown in Fig. 15.

4. Similar to the background

The images that closely similar to the background, as

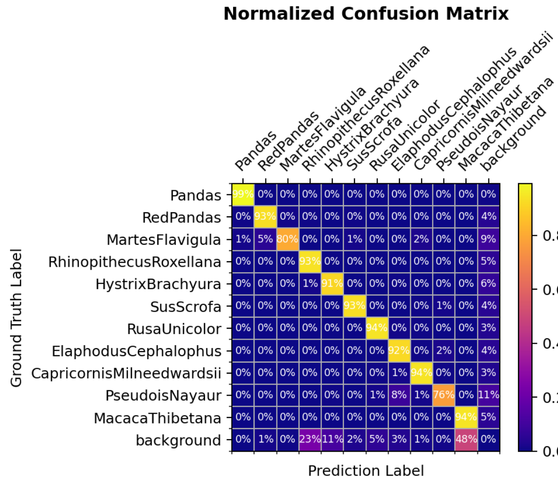

Figure 7: Queryinst on wild subset.

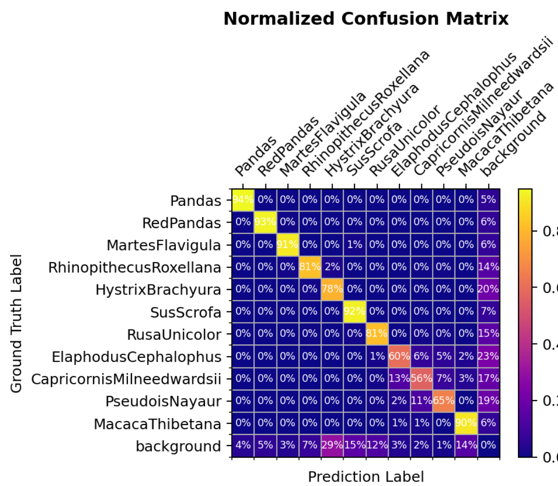

Figure 8: Queryinst on web subset.

shown in Fig. 16.

5. Other situations

Other situations includes the ratio of animals is too large; large GT, but small DT; local prediction of one animal into two animals; only one bounding box was predicted for multiple animals; false detection and so on, as shown in Fig. 17.

3.2. Instance segmentation

We also analyze the visualization results of wild animal instance segmentation and found that the outline of animals are more complex than bounding boxes, especially for animals with thin and long limbs, long horns on their heads, long and thin tails, and needle-like hair on their bodies. Issues such as occlusion, similar backgrounds with the body, and animals being cut into several parts or having a small

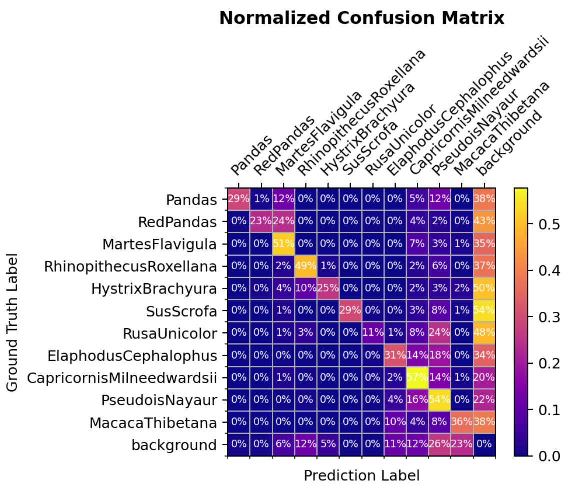

Figure 9: Queryinst on cross subset.

Figure 10: Remote and small animals.

Figure 11: Animals interaction.

Figure 12: Environmental occlusion.

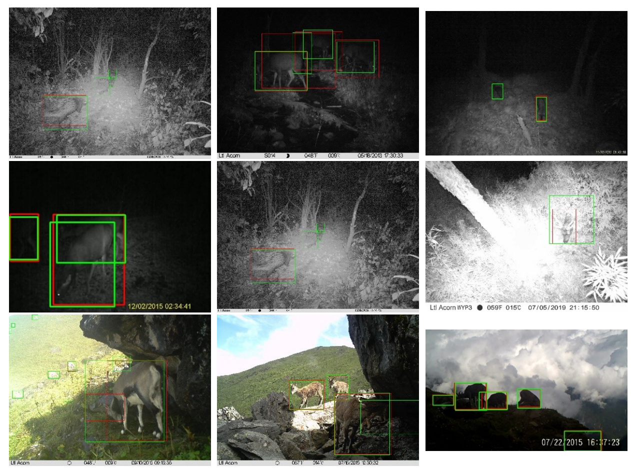

Figure 13: Environmental occlusion.

Figure 14: Foggy weather.

proportion of pixels in small targets also negatively affect the performance of instance segmentation. Additionally, we found instances of segmentation accuracy but classification

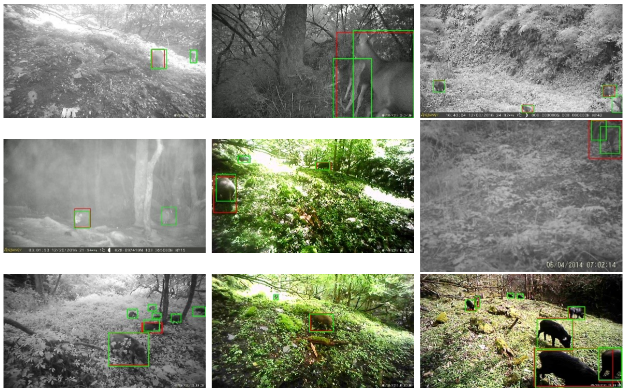

Figure 15: Lower pixel images.

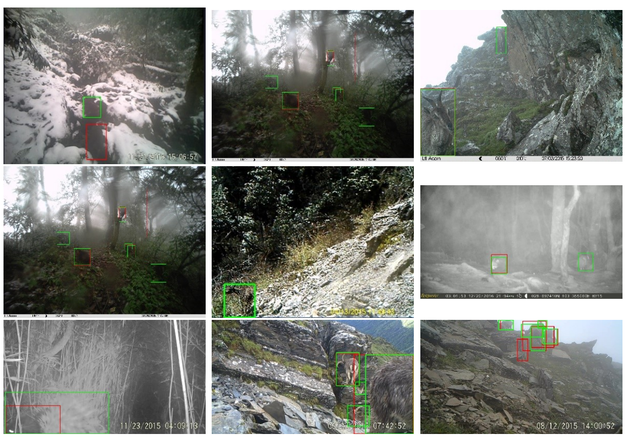

Figure 16: Similar animals and backgrounds.

Figure 17: Other situations.

errors. The visualization of these questions are as follows. Note: (Green mask: GT, dark blue mask: DT)

1. Complex mask outline

Outline edges of animals are more complex than their bounding box. Especially animals with slender limbs and tails, long horns on the top of the head and needle-like hair on the body, as shown in Fig. 18.

Figure 18: Complex mask outline.

2. Occlusion problem

(1) The first type is animal interactive occlusion, resulting in model failure to accurately segment overlapping animal outlines, as shown in Fig. 19.

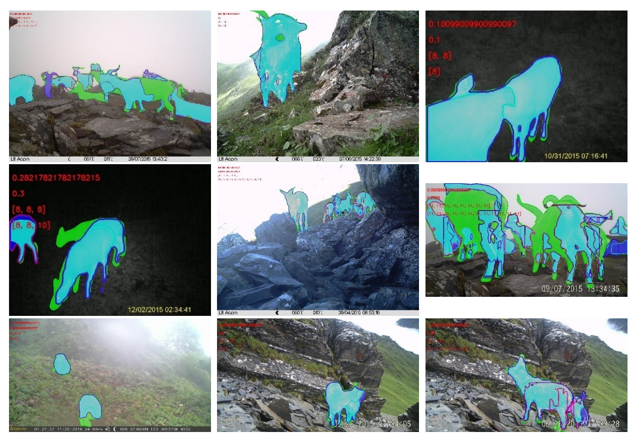

Figure 19: Animals interactive occlusion.

(2) The second type is the blurred outline of animals caused by environmental occlusion, as shown in Fig. 20.

3. Wrong position

The background is similar to the animal body, but the positioning animal is wrong, as shown in Fig. 21.

4. Truncated mask

Animal with multiple masks is incomplete segmentation mask, as shown in Fig. 22.

5. Smaller pixel area ratio

Small objects have a smaller pixel area ratio, as shown in Fig. 23.

Figure 20: Blurred outline of environmental occlusion.

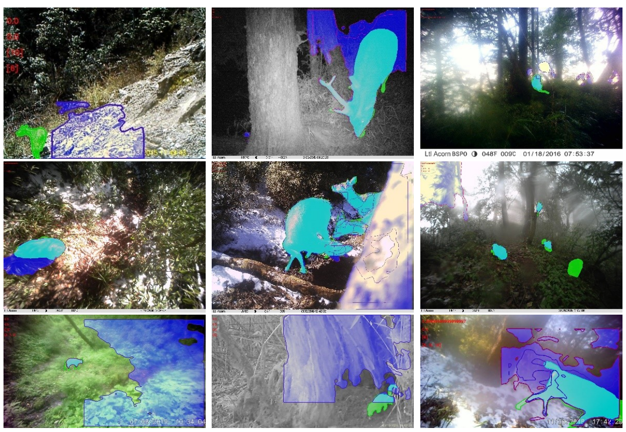

Figure 21: Animals positioning is wrong.

Figure 22: The truncated mask.

6. Wrong classification

Segmentation is accurate, but classification is wrong, as shown in Fig. 24.

Figure 23: Animals truncation.

Figure 24: Wrong classification.

3.3. Pose estimation

We also analyze the visualization results of wild animal pose estimation. We found that issues such as animal interaction or environmental occlusion, crowded scenes, boundary truncation, image blurring, low resolution, small targets, and short animal limbs all negatively affect the performance of pose estimation. The visualization of these questions are as follows.

Note: (Green: GT keypoints, red: DT keypoints)

1. Animals interaction or environmental occlusion

Animals interaction or environmental occlusion, such as climbing, grooming and other interaction behaviors, and grass trunk occlusion, as shown in Fig. 25.

2. Crowded scenes

The animals crowded together, as shown in Fig. 26.

3. Animal body truncation

The animal's skeletal keypoints are not in full view of the camera, as shown in Fig. 27.

4. Blurred and low resolution

The animal is in the blurred and low resolution image, as shown in Fig. 28.

5. Small objects

Figure 25: Animals interaction or environmental occlusion.

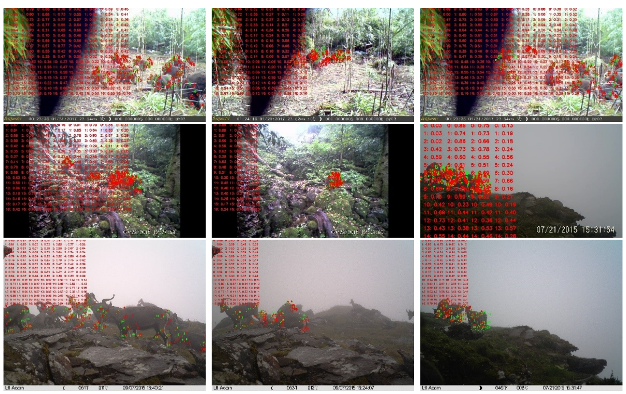

Figure 26: Crowded scenes.

Figure 27: Animal body truncation.

Small animal carry dense skeletal keypoints, which converge into a pile close together, as shown in Fig. 29.

6. Back or tail view

From the back or tail view, the animal's head and limb skeleton key points are hidden, as shown in Fig. 30.

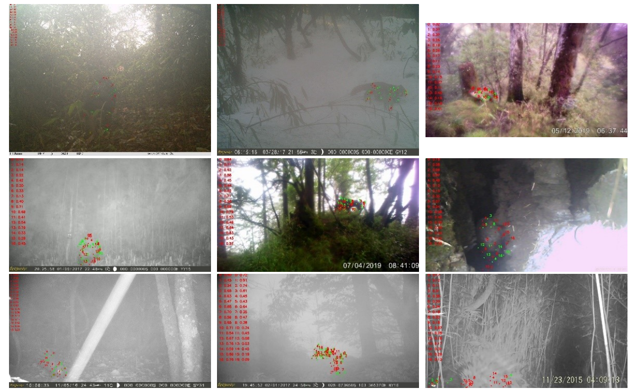

Figure 28: Blurred and low resolution image.

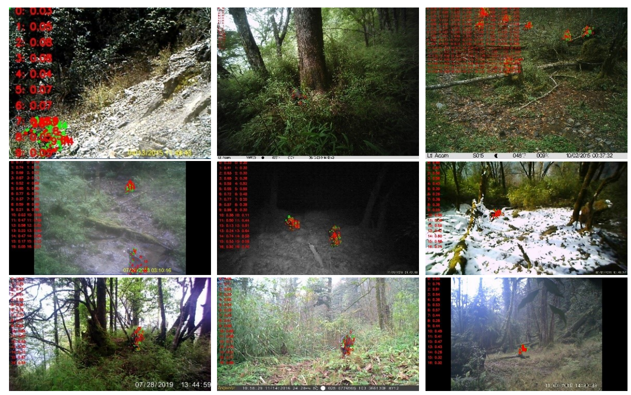

Figure 29: Small objects.

Figure 30: Back or tail view.

3.4. Action recognition

Finally, we analyze the visualization results of wild animal behavior recognition. We found that issues such as diverse scenes, seasons, weather, day and night variations, and differences in species' movements all negatively affect the performance of behavior recognition. The visualization of these questions are as follows.

1. Rich scenes

The following scenes include water sources, branches, grass, ridges, rock crevasses, nests, animal trails, ice, clouds, etc., as shown in Fig. 31.

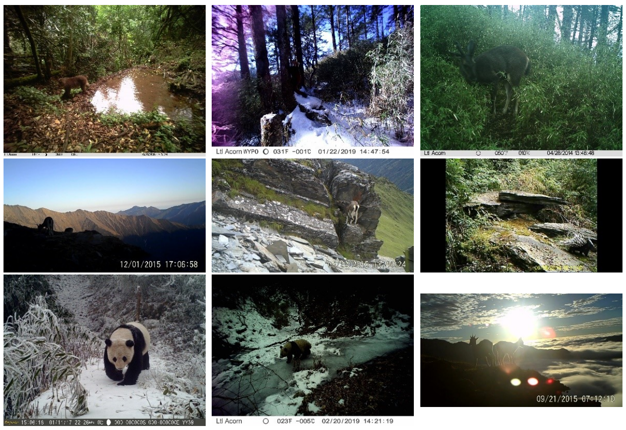

Figure 31: Rich scenes.

2. Variable seasons and weather

The seasons, weather, day and night are variable, as shown in Fig. 32. The first row shows three ecological seasons: spring, summer-autumn and winter. The second row is three weather days: fog, rain, sunny. The third row is morning, afternoon and late night.

Figure 32: The seasons, weather, day and night.

3. Interspecific behavioral differences

Different species have different interspecific behaviors, as shown in Fig. 33. The first row shows Sambar (*Rusa Unicolor*) drink water behavior, The second row shows Red Panda (*Ailurus Fulgens*) drink water behavior. The third row shows Tibetan Macaque (*Macaca Thibetana*) drink water behavior.

Figure 33: Interspecific behavioral differences.

In conclusion, visual tasks in outdoor scenes are challenging, and require more detailed analysis and optimization based on the characteristics of each task. Through the above visualization results, we can better understand the reasons for various issues and provide guidance for further model improvement.