Recognition in Terra Incognita: Supplementary Material

1 Additional Experiments

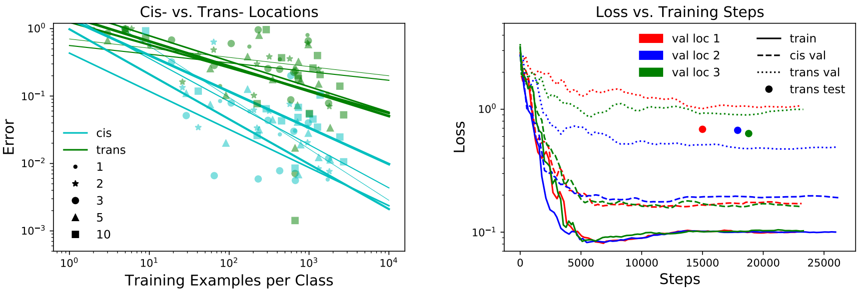

Fig. 1. Generalization metrics are robust to N. locations and to validation. Both plots are based on bounding box classification. **(Left)** Error per class vs. number of training examples (best-fit line width denotes number of training locations in $1, 2, 3, 5, 10$). Trans performance is stable for N locations with $2 < N \leq 10$. We chose 10 training locations to study generalization behaviors while providing maximal data for experimentation. **(Right)** Loss curves using different locations for trans-validation. The test loss for the selected model for each validation set remains stable, implying that the choice of validation location does not greatly impact trans test performance.

1.1 Varying the amount of training data per location

We chose to use a small number of locations as this is a key variable of the generalization problem. In the limit, we would study the behavior of models trained on a single location with "unlimited" training data. We did not have access to such a dataset, and therefore used 10 training locations in order to have a sufficient number of training examples. To verify whether 10 training locations would yield significantly different results than 1 training location, we ran our bounding box experiments with a quarter, half, and all of the images available per training location, and saw trans test accuracies of 80.6%, 83.0%, and 83.4% respectively. This implies that *increasing the number of images per location would not solve the generalization problem.*

1.2 Varying the number of training locations

As an additional control, we experimented with varying the number of training locations (see Fig. 1(Left)), and find that trans performance is stable as the number of training locations is increased beyond 2. Thus, we are confident that our dataset is adequate to measure generalization ability. We expect the generalization gap to narrow with $N >> 10$, but as the number of training locations increases the focus of the experiment shifts. We want to provide a test bed to specifically study generalization when provided with few training locations.

1.3 Varying the validation location

To analyze the effect of the validation split, we repeated our experiments with 2 other validation locations (see Fig. 1(Right)). We find that test performance is relatively stable regardless of the validation split. Fig. 1(Right) also shows training and validation curves for the three different validation experiments.

2 Data Format

We chose to use an adapted version of the JSON format used by the COCO dataset with additional camera trap-specific fields, which we call COCO-CameraTraps. The format can be seen in Fig. 2.

We added several fields for each image in order to specify camera-trap specific information. These fields include a location id, a sequence id, the number of frames in that sequence, and the frame number of the individual image. Note that not all cameras take sequences of images at a single trigger, so for some images the number of frames in the associated sequence will be one.

All data can be accessed at https://beerys.github.io/CaltechCameraTraps/.

```
{
  "info" : info,
  "images" : [image],
  "categories" : [category],
  "annotations" : [annotation]
}

info{
  "year" : int,
  "version" : str,
  "description" : str,
  "contributor" : str
  "date_created" : datetime
}

image{
  "id" : str,
  "width" : int,
  "height" : int,
  "file_name" : str,
  "rights_holder" : str,
  "location": int,
  "datetime": datetime,
  "seq_id": str,
  "seq_num_frames": int,
  "frame_num": int
}

category{
  "id" : int,
  "name" : str
}

annotation{
  "id" : str,
  "image_id" : str,
  "category_id" : int,
  "bbox": [x,y,width,height]
}
```

Fig. 2. COCO-CameraTraps data format