iLab-2M

Toy Leksut

Figure 1: The iLab20M dataset provides toy vehicle images from 15 categories in various viewpoints, lighting conditions, and backgrounds. Cropped images of sample object instances for each category are shown above. Sample images demonstrate rich intra-class variations in shape, size, and color.

1 The iLab-2M Dataset

1.1 iLab-20M

The iLab-20M dataset [1] is a large-scale controlled, parametric dataset of toy vehicle objects under variations of viewpoint, lighting, and background. The dataset is produced by placing a physical object on a turntable and using multiple cameras located on a semicircular arc over the table. Toy vehicle instances are annotated into the following 15 categories: boat, bus, car, equipment, f1car, helicopter, military, monster truck, pickup truck, plane, semi truck, tank, train, UFO, and van. Each category consists of 25 to 160 object instances, and the whole dataset contains 718 object instances. For each object instance, the dataset provides 88 different viewpoints from 11 camera azimuth angles and 8 turntable rotation angles, 5 lighting conditions, 3 camera focus settings, and 14 – 40 background images. This enhances the number of training images for a single object to 1320 images per object instance per background. The total number of images is approximately 22 million.

All parameters are listed as follows:

- $\mathbf{y} = \{$*boat, bus, car, equip, f1car, heli, mil, monster, pickup, plane, semi, tank, train, ufo, van*$\}$ is the set of object categories.

- $\mathbf{i} = \{i_0, ..., i_{717}\}$ is the set of object instance identification numbers for a total of 718 instances.

- $\mathbf{c} = \{c_0, ..., c_{10}\}$ specifies eleven cameras which provide eleven viewpoints varying in azimuth angle.

- $\mathbf{r} = \{r_0, ..., r_7\}$ specifies eight rotation angles around the center of the turntable.

- $\mathbf{l} = \{l_0, ..., l_4\}$ specifies five lighting conditions produced by four light bulbs. The first four conditions are from one of the four lights on, and the fifth one is from when all of the lights are on.

- $\mathbf{f} = \{f_0, f_1, f_2\}$ specifies three focus settings.

- $\mathbf{b} = \{b_0, ..., b_{131}\}$ is the set of background images. There are a total of 125 printouts of satellite images and 7 solid-color backgrounds. Every object instance comes with at least 14 relevant backgrounds e.g. cars on roads, boats on water, and trains on rails.

In summary, each object instance has 1,320 images per background resulting from 11 cameras $\times$ 8 rotations $\times$ 5 lightings $\times$ 3 focuses, and the number of backgrounds for each instance varies from 14 to 40. Sample cropped images of object instances from all categories are shown in Figure 1.

Note that the number of instances per category is unbalanced and ranges from 25 to 160. For example, the car category has 160 instances, whereas the boat category has 26 instances. This results in approximately 4.8 million more car images compared to boat images.

This dataset is freely available upon request from the authors and upon shipping of two 12TB blank hard drives.

1.2 iLab-80M

The iLab-80M is an augmented set of the iLab-20M adding random crops and scales. The augmentation ensures that the number of images per category is well balanced, resulting in 5.5 million images per category and a total of 82.7 million images for the whole set. The iLab-80M data distributions are demonstrated in Figure 2.

In addition, the original 960×720 images are cropped around each object and rescaled to 256×256. This dataset is also freely available from the authors upon request, and fits on a single 8TB hard drive thanks to the smaller image size.

1.3 iLab-2M

: From iLab-80M, we generate our own subset for experiments conducted in this work. Our focus variation is pose, therefore we select data with pose variations in mind and keep other visual variations, in this case lighting and focus, constant.
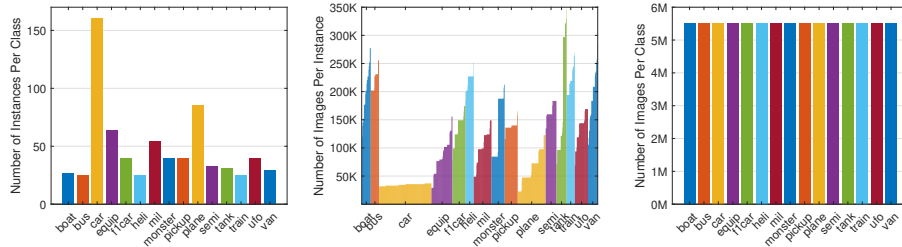
Figure 2: The augmented iLab-80M data distributions. Originally in the iLab-20M dataset, the number of images per instance are the same for all instances, which results in data imbalance in terms of number of images per category. By augmenting iLab-20M using random crops and scales, the number of images per category are balanced in iLab-80M as shown in the right chart. (Left) Number of object instances per object category remains the same before and after augmentation. (Middle) Number of images per instance after augmentation. Object instance bars are colored with respect to category. (Right) Number of images per object category after augmentation.

With this criterion, we present the iLab-2M, a subset of the iLab-80M with the following selected parameters:

- five cameras $c_i \in \{c_0, c_2, c_4, c_7, c_9\}$

- six rotations $r_i \in \{r_1, r_2, r_3, r_4, r_6, r_7\}$

- one lighting condition $l_i \in \{l_0\}$

- one focus value $f_i \in \{f_2\}$

- all categories $y_i \in \mathbf{y}$

- all instances $i_i \in \mathbf{i}$

We sample a total of 1,751,719 images (1.7M) and partition them into 70% training (1.2M), 15% validation (270K), and 15% test (270K) splits. Figure 4 shows the category distribution for each split. Sets of object instances in training, validation, and test splits are non-overlapping. We include objects from all 15 categories and a combination of 5 cameras and 6 rotations which results in 30 different poses. The dataset is available to download from `http://ilab.usc.edu/ilab2m/iLab-2M.tar.gz`.

1.4   iLab-2M-Light

We create an extension of the iLab-2M dataset to include all five lighting conditions $l_i \in \{l_0, l_1, l_2, l_3, l_4\}$. Other parameters remain the same as in iLab-2M. We sample a total of 1,999,743 images and partition them into 70% training
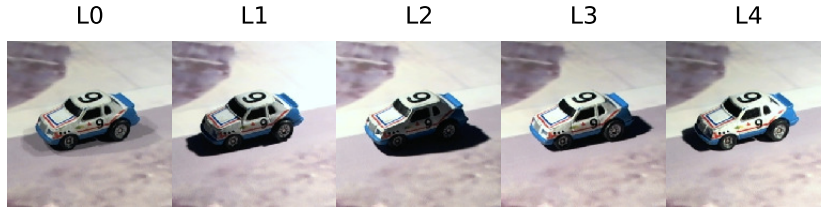
Figure 3: In addition to pose, lighting condition is a controlled visual variation provided by the iLab-20M dataset. Shown above is an example of five available lighting conditions.
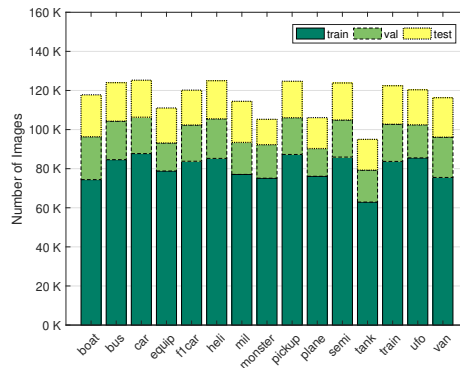

Figure 4: The distribution of images per object category for iLab-2M training, validation, and test splits. The number of training images is 1.2 millions. Validation and test sets each contains 270K images. Object instances in all three sets are mutually exclusive.

(1.36M), 15% validation (316K), and 15% test (316K) splits. The iLab-2M-Light dataset is available to download from `http://ilab.usc.edu/ilab2m/iLab-2M-Light.tar.gz`.

References

[1] A. Borji, S. Izadi, and L. Itti, "ilab-20m: A large-scale controlled object dataset to investigate deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2221–2230.