SSN: Shape Signature Networks for Multi-class Object Detection from Point Clouds

Anonymous ECCV submission

Paper ID 5131

1 Network Architecture

In this section, we describe details of the proposed shape-aware grouping heads. Since the categories are different in Lyft [1] and nuScenes [2] datasets, we thus apply different shape-aware grouping heads for them. We first describe the basic block used in the heads. Then we give the specific settings, respectively. Finally, we also give the **code** for the clear explanation.

The basic block consists of Convolution, BatchNorm and Relu activation. In Lyft dataset, {"pedestrian" and "animal"}, {"bicycle" and "motorcycle"}, {"car" and "emergency-vehicle"}, {"bus", "other-vehicle", "truck"} are four heads, which are grouped based on the shape. More specifically, {64, 64}, {64, 64}, {128, 128, 64(stride=2), 64}, {256, 256, 128(stride=2), 128, 64(stride=2), 64} are the size of convolutional kernel. We also apply the stride=2 in the convolution layer to perform the downsample.

In nuScenes dataset, {"pedestrian" and "traffic cone"}, {"bicycle" , "motorcycle", "barrier" }, {"car"}, {"bus", "trailer", "truck" and "construction vehicle"} are four heads. The settings of convolutional kernel are {64, 64, 64}, {64, 64, 64}, {128, 128, 128, 64(stride=2), 64, 64}, {256, 256, 128(stride=2), 128, 128, 64(stride=2), 64, 64}, respectively.

Here, we also give the code example to clearly show the head. We show the {"pedestrian" and "animal"} head in Lyft as following. Note that our full codebase will be available upon publication.

```
ped_ani_net = nn.Sequential(
            nn.Conv2d(num_filters, 64, 3, bias=False,
            padding=1, dilation=1),
            nn.BatchNorm2d(64, eps=0.001, momentum=0.01),
            nn.ReLU(),
            nn.Conv2d(64, 64, 3, bias=False, padding=1),
            nn.BatchNorm2d(64, eps=0.001, momentum=0.01),
            nn.ReLU(),
            )
```

2 Evaluation Metrics

In this section, we give the details of evaluation metrics in nuScenes dataset. We first give the definitions of Average Precision metrics, including Average

Translation Error (ATE), Average Scale Error (ASE), Average Orientation Error (AOE), Average Velocity Error (AVE), and Average Attribute Error (AAE). Specifically, Average Translation Error (ATE) is the Euclidean center distance in 2D (units in meters). Average Scale Error (ASE) is the 3D IOU after aligning orientation and translation ($1 - \mathrm{IOU}$). Average Orientation Error (AOE) is the smallest yaw angle difference between prediction and ground-truth (radians). Average Velocity Error (AVE) is the absolute velocity error as the L2 norm of the velocity differences in 2D (m/s). Average Attribute Error (AAE) is defined as 1 minus attribute classification accuracy ($1 - \mathrm{acc}$).

Then the NDS is shown as following:

$$\mathrm{NDS} = \frac{1}{10}[5\ \mathrm{mAP} + \sum_{\mathrm{mTP} \in \mathbb{TP}} (1 - \min(1, \mathrm{mTP}))] \tag{1}$$

It is worthy to note that the mAP metric in Lyft defines a match by thresholding the 2D center distance d on the ground plane instead of intersection over union ($d = \{\ 0.5,\ 1,\ 2,\ 4\ \}$).

For Lyft dataset, the results are evaluated on the mean average precision at different intersection over union (IoU) thresholds.

$$IoU(A, B) = \frac{A \cap B}{A \cup B}. \tag{2}$$

The metric sweeps over a range of IoU thresholds, at each point calculating an average precision value. The threshold values range from 0.5 to 0.95 with a step size of 0.05: (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95). Hence, the metric of mAP-3D is much difficult than mAP in nuScenes and Kitti.

3 Submissions on Official Websites

We give the feedback from official evaluation websites to verify the results of our experiments in the attachment, including the json file from nuScenes evaluation server and the leaderboard from Lyft evaluation website.

References

1. level 5 dataset, L.: https://level5.lyft.com/dataset/, https://level5.lyft.com/dataset/
2. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. arXiv preprint arXiv:1903.11027 (2019)